La naturalización de las emociones:
anotaciones a partir de Wittgenstein

Juan Raúl Loaiza Arias



Opera Prima
Escuela de Ciencias Humanas



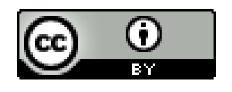


La naturalización de las emociones: anotaciones a partir de Wittgenstein

Juan Raúl Loaiza Arias

SciELO Books / SciELO Livros / SciELO Libros

LOAIZA ARIAS, J. R. *La naturalización de las emociones*: anotaciones a partir de Wittgenstein [online]. Bogotá: Editorial Universidad del Rosario, 2016, 174 p. Opera prima collection. ISBN: 978-958-738-724-7. Available from: https://books.scielo.org/id/drng3. https://doi.org/10.12804/op9789587387247.



All the contents of this work, except where otherwise noted, is licensed under a Creative Commons Attribution 4.0 International license.

Todo o conteúdo deste trabalho, exceto quando houver ressalva, é publicado sob a licença Creative Commons Atribição 4.0.

Todo el contenido de esta obra, excepto donde se indique lo contrario, está bajo licencia de la licencia Creative Commons Reconocimento 4.0.


Universidad del
Rosario

LA NATURALIZACIÓN DE LAS EMOCIONES:

anotaciones a partir de Wittgenstein

La naturalización de las emociones: anotaciones a partir de Wittgenstein

Resumen

En la literatura sobre las emociones una de las teorías con mayor fuerza es la llamada "teoría James-Lange". En esta obra se intenta hacer una crítica a dicha teoría a partir de algunas observaciones de Wittgenstein sobre el uso de conceptos psicológicos, sacando a la luz dos confusiones gramaticales que surgen en ella. Para ello, se construye primero la categoría de "programa de naturalización de las emociones" que recoge las teorías del Descartes, James y Prinz, siguiendo la metodología de Lakatos. Luego, se identifica como problema central el de la naturalización de la intencionalidad. Para luego exponer algunas herramientas de Wittgenstein para estudiar la gramática de la pregunta por el objeto y la intencionalidad las emociones, mostrando que las respuestas del programa de naturalización no son satisfactorias y no respetan las reglas de ciertos usos del lenguaje.

Palabras clave: Emociones y sentimientos, filosofía de la psicología, intencionalidad - filosofía, Ludwig Josef Johann Wittgenstein, 1889-1951, crítica e interpretación.

The naturalization of emotion: notes after Wittgenstein

Abstract

One of the most influential approaches in the literature on emotion is the James-Lange theory. This text suggests some criticisms of the theory through the use of certain observations made by Wittgenstein regarding grammatical confusions that it engenders. The author constructs the category of "naturalization of emotions research", following the methodology of Lakatos to consider the theories of emotion proposed by Descartes, James, and Prinz. The naturalization of intentionality is identified as the central problem and some of Wittgenstein's recommendations for studying the grammar of questions for objects and the intentionality of emotions are considered, demonstrating that the answers provided by naturalization of emotions research are unsatisfactory and do not respect the rules of certain uses of language.

Keywords: Emotion and Sentiment, Philosophy of Psychology, Intentionality - Philosophy, Ludwig Josef Johann Wittgenstein, 1889-1951, Criticism and Interpretation.

Para citar este libro
Loaiza Arias, Juan Raúl, *La naturalización de las emociones: anotaciones a partir de Wittgenstein.* Bogotá: Editorial Universidad del Rosario, 2016.
DOI: dx.doi.org/10.12804/op9789587387247

LA NATURALIZACIÓN DE LAS EMOCIONES:

anotaciones a partir de Wittgenstein

JUAN RAÚL LOAIZA ARIAS

Loaiza Arias, Juan Raúl
La naturalización de las emociones: anotaciones a partir de Wittgenstein / Juan Raúl Loaiza Arias – Bogotá: Editorial Universidad del Rosario, Facultad de Ciencias Humanas, 2016.
174 páginas. (Colección Opera Prima)
Incluye referencias bibliográficas

ISBN: 978-958-738-723-0 (impreso)
ISBN: 978-958-738-724-7 (digital)
DOI: dx.doi.org/10.12804/op9789587387247

Emociones y sentimientos (Filosofía) / Filosofía de la psicología / Intencionalidad (Filosofía) / Wittgenstein, Ludwig Josef Johann, 1889-1951 - Crítica e interpretación, etc. / I. Universidad del Rosario. Facultad de Ciencias Humanas / II. Título / III. Serie.

128.37 SCDD 20

Catalogación en la fuente – Universidad del Rosario. Biblioteca

JDA abril 14 de 2016

Hecho el depósito legal que marca el Decreto 460 de 1995



Colección Opera Prima

© Editorial Universidad del Rosario
© Universidad del Rosario
Facultad de Ciencias Humanas
© Juan Raúl Loaiza Arias

Editorial Universidad del Rosario
Carrera 7 Nº 12B-41, oficina 501
Teléfono 297 02 00, ext. 3113
editorial.urosario.edu.co

Primera edición: Bogotá D.C., mayo de 2016

ISBN: 978-958-738-723-0 (impreso)
ISBN: 978-958-738-724-7 (digital)
DOI: dx.doi.org/10.12804/op9789587387247

Coordinación editorial:
Editorial Universidad del Rosario
Corrección de estilo: Claudia Ríos
Diseño de cubierta y diagramación:
Precolombi EU-David Reyes
Impresión: Digiprint

Impreso y hecho en Colombia
Printed and made in Colombia

Fecha de evaluación: 13 de agosto de 2014
Fecha de aceptación: 29 de marzo de 2016

Todos los derechos reservados. Esta obra no puede ser reproducida sin el permiso previo por escrito de la Editorial Universidad del Rosario.

Contenido

Agradecimientos

Primero que todo, quiero agradecer al profesor Carlos Cardona, quien ha leído y comentado cada palabra de este trabajo. A él no solo debo agradecerle su atención con este proyecto, sino todo el ánimo y todas las oportunidades que me ha dado a lo largo de mis años de estudio. De él aprendí lo que es leer, escribir y estudiar con paciencia y cuidado, a ir paso a paso, revisando siempre lo que se ha hecho y progresando poco a poco hasta lograr lo propuesto. Si algo de eso ha quedado, espero que esté reflejado en las páginas que vienen a continuación.

Segundo, me gustaría agradecer al profesor Carlos Patarroyo por su dedicación y atención. Este trabajo, a veces directamente y a veces indirectamente, es producto de varias lecciones suyas. En especial, es un intento por tomar posición, por apropiarme de algunas ideas y exponerlas con algún sentido. Dicho de otra manera, este trabajo es la búsqueda —quizás torpe— de una voz. Esta fue la mejor lección que él pudo haberme dejado.

Finalmente, agradezco a quienes han tenido la paciencia y el estoicismo para leer y comentar este trabajo. En especial, gracias a Fabio Fang por todos sus comentarios, tanto fuera como dentro del texto. También agradezco a Alejandro González, Ernesto Navarro, Juliana Gutiérrez,

Dimitri Coelho, Robin Löhr, Brad Scott y Dan Cook, quienes ayudaron a concretar, aclarar y sintetizar varias de las ideas aquí expuestas. Por último, agradezco a mis jurados, la Dra. Magdalena Holguín y el Dr. Miguel Ángel Pérez, por sus observaciones y críticas.

Abreviaturas

LPA	Descartes, R. (1649/2010). *Las pasiones del alma*
PP	James, W. (1890/1950). *The principles of psychology*
IF	Wittgenstein, L. (1953/2003). *Investigaciones filosóficas*
CA	Wittgenstein, L. (1958/2009). *Cuaderno azul*
OFP	Wittgenstein, L. (1980/2007). *Observaciones sobre la filosofía de la psicología*

Introducción

Sentimos que aun cuando todas las posibles cuestiones científicas hayan recibido respuesta, nuestros problemas vitales todavía no se han rozado en lo más mínimo. Por supuesto que entonces ya no queda pregunta alguna; y esto es precisamente la respuesta.

Wittgenstein, 1921/1999, §6.52

No son pocas las veces en las que nos topamos con ese aparentemente extraño fenómeno que son nuestras emociones. A menudo usamos expresiones como "las emociones no son racionales" o "esta persona es muy emocional, pero debería ser más racional". Esto sugiere que las emociones, por un lado, están íntimamente ligadas a nuestras vidas y que hacen parte de nuestra cotidianidad, pero además, que ellas nos evocan preguntas que no parecen fáciles de responder y que incluso a veces pensaríamos que están por fuera del alcance de la razón.

Sin embargo, hay formas claras en las que, en esa cotidianidad, damos razones y aclaramos por qué sentimos emociones. Pensemos en un caso sencillo: imaginemos a una persona quien siente tristeza por la muerte de algún familiar, o que siente alegría por haber terminado de escribir un

libro. Esta persona ciertamente diría que tiene una buena razón para sentir su emoción, esto es, que su tristeza o su alegría están bien fundadas. Si en cambio la persona sintiese alegría por la muerte de su familiar, o tristeza por haber terminado de escribir el libro, intentaríamos buscar una manera de justificarlo aduciendo a otras razones de la persona. Por ejemplo, podríamos decir que la persona entonces es egoísta y quiere quedarse con la herencia de su familiar fallecido o que se ha apegado mucho a la tarea de escribir el libro y entonces resulta comprensible que sienta tristeza al terminarla. Todo esto indica que nuestras razones, de algún modo, entran a jugar un papel importante en la manera en que hablamos y pensamos nuestras emociones. En un lenguaje más técnico, diríamos que creencias y deseos tienen alguna relación con nuestras emociones.

Esto, no obstante, nos lleva a un clásico problema filosófico. Supongamos por un momento que somos científicos intentando estudiar la manera en que las personas sienten emociones. ¿Qué buscaríamos? ¿Qué examinaríamos? En un famoso artículo publicado en 1884, William James sugería que nuestras emociones no eran más que la percepción de reacciones en nuestro cuerpo (James, 1884). Siguiendo a James, tendríamos que examinar entonces las respuestas cerebrales o los cambios de ritmo cardiaco de las personas, pero esto parece no tener relación con cosas como "la muerte de un familiar" o "terminar de escribir un libro". Si llegase un paciente y yo examinara únicamente su cuerpo, sin hacerle ningún tipo de pregunta sobre sus emociones,

a lo sumo podría decir que siente tristeza por algo, pero difícilmente sabría por qué.

En las discusiones filosóficas este problema —o al menos una de sus versiones— es conocido como el problema de la intencionalidad. El término 'intencionalidad' se usa para referirse a un rasgo de nuestros estados mentales que es el de tener un objeto, por ejemplo, cuando veo un árbol, mi estado mental es el de "ver", y el objeto de mi visión es el árbol. Esto parece también aplicar a las emociones: cuando una persona siente tristeza por la muerte de un familiar su estado mental es el de la tristeza y el objeto de su tristeza es la muerte del familiar. El problema de la intencionalidad surge cuando intentamos describir nuestra vida mental a partir de sus correlatos físicos. Cuando el científico estudia las activaciones cerebrales o el cambio de ritmo cardiaco, nada de eso parece decirle cuál es el objeto de la emoción de su paciente. Volviendo al ejemplo anterior, si el científico no hace pregunta alguna, parece difícil saber que el objeto de la tristeza de su paciente es la muerte de su familiar.

Dentro de los debates en filosofía sobre las emociones este problema no ha sido menor. Buena parte de las teorías de las emociones que pueden verse en el debate parten de una u otra manera de atacar esta dificultad. Como ya mencioné, el problema parece surgir al intentar dar una descripción de nuestros estados mentales en términos físicos, en el caso de las emociones, al intentar describirlas únicamente como la percepción de cambios en el cuerpo, siguiendo la idea de James. Algunos han optado por abandonar la tesis de James

y elaborar teorías de las emociones como juicios de valor y así distinguir las emociones de las sensaciones (reacciones corporales) que las acompañan (Nussbaum, 2001; Solomon, 1973). A estas teorías se les llama en la literatura 'teorías cognitivistas'. Otros han intentado hacer teorías híbridas que rescaten algunos de los aspectos fisiológicos a los que James apunta a la vez que incluyen algunas ideas de las teorías cognitivistas (Helm, 2009). También hay quienes sostienen que en realidad lo que llamamos 'emociones' son cosas diferentes en el mundo y que entonces no es posible hacer una teoría general sobre ellas (Griffiths, 1997).

Sin embargo, hay un movimiento interesante que queda disponible, que es el que me interesa examinar en este texto. Jesse Prinz (2004) ha elaborado una teoría naturalista de las emociones que rescate la teoría de James, pero que logra superar el problema de la intencionalidad. En otras palabas, Prinz desarrolla una teoría de las emociones que, manteniendo la idea de que ellas no son más que la percepción de reacciones corporales, explique cómo ellas refieren a cosas como la muerte de un familiar o terminar de escribir un libro. Esta maniobra exige, por supuesto, dar una teoría de cómo los correlatos físicos pueden decirnos algo sobre los objetos de nuestros estados mentales, cómo pasar de lo físico a lo intencional; en síntesis, esta maniobra exige naturalizar la intencionalidad. Con ello tendríamos a su vez una teoría naturalizada de las emociones.

No obstante, me gustaría poner en entredicho el alcance de este tipo de teoría como teoría filosófica. En este trabajo intentaré construir algunas críticas a las teorías de natura-

lización de las emociones. Para ello usaré algunas de las anotaciones que realizó Wittgenstein a lo largo de su obra sobre varios temas relativos a la psicología. Es mucho lo que se ha escrito sobre este tema y quizás aquí solo logremos cubrir una pequeña parte del abanico de posibilidades que ofrece este autor. Sin embargo, basta con dar algunas luces sobre cómo un proyecto que intente naturalizar las emociones enfrenta dificultades conceptuales.

Si lo que nos ocupa es un problema filosófico, entonces lo primero que hemos de hacer, parafraseando a Wittgenstein (IF, §123), es sumergirnos en el atolladero. Para ello, exploraremos en qué consiste la naturalización de las emociones. En el primer capítulo, reconstruiremos las teorías de René Descartes, William James, y Jesse Prinz. Estas se han considerado en el canon como las principales teorías naturalistas sobre las emociones, en especial las últimas dos.[1] Todas ellas comparten la tesis general que mencioné anteriormente, según la cual las emociones son la percepción de reacciones corporales. Para ver de manera general la continuidad entre estas teorías, usaré la matriz historiográfica que ofrece Imre Lakatos a propósito de los programas de investigación científica. Ver en Descartes, James y Prinz un programa de investigación a la manera de Lakatos permitirá hablar en abstracto de un "programa de naturalización de las emociones".

---

[1] Prinz, por ejemplo, considera que él está continuando la tradición que comenzó Descartes y que pasa por James (2004b, p. 242). Sin embargo, pese a que la teoría puede rastrearse hasta Descartes, esta se conoce en la literatura como la teoría James-Lange, en nombre del filósofo norteamericano y el fisiólogo Carl Lange.

Ahora bien, una vez identificado el programa de naturalización de las emociones, procederemos a sumergirnos aún más en el atolladero y entraremos en un problema particular de este programa: la naturalización de la intencionalidad. Esto ocupará el segundo capítulo, donde examinaremos en qué consiste esta empresa y cómo el programa de naturalización de las emociones ha respondido a la exigencia de naturalizar la intención.[2] Con ello intentaremos ver en la naturalización de la intencionalidad un paso importante en la naturalización de las emociones y cómo el programa progresó paulatinamente hasta alcanzar una teoría sobre la intencionalidad.

Luego de esto, estaremos en mejor posición de ofrecer las críticas mencionadas al programa de naturalización de las emociones. Para este propósito, haremos un despliegue de algunas herramientas que ofrece Wittgenstein para estudiar conceptos psicológicos. Sin embargo, antes de ello es menester hacer una breve digresión sobre la filosofía wittgensteiniana. Así las cosas, en el tercer capítulo pasaremos de una discusión científica sobre las emociones y la intencionalidad a una discusión gramatical. Una vez aclarada la naturaleza de la investigación conceptual de Wittgenstein, podemos desplegar las herramientas mencionadas y aplicarlas al programa de naturalización de las emociones. En particular, dirigiremos nuestras críticas al problema

---

[2] Si bien es discutible si las emociones son intencionales, en este texto me pliego a una lectura de las teorías mencionadas que muestra que ellas aceptan el carácter intencional de ellas.

desarrollado en el segundo capítulo, i.e., la naturalización de la intencionalidad. Con esto pretendo mostrar que el programa de naturalización de las emociones, al contener en su núcleo el problema de la naturalización de la intencionalidad, encierra varias confusiones conceptuales.

Si bien son varias las observaciones positivas que Wittgenstein dedica a elucidar el concepto de emoción (tales como detectar en las emociones una duración genuina o que estas no estén localizadas), debo advertir de antemano que ellas no serán objeto de este trabajo. Las anotaciones a propósito de las emociones son muy variadas y su aclaración, por sí sola, sería motivo de un trabajo entero. Por lo pronto, nos centraremos en usar algunas herramientas para comenzar a disolver una ilusión filosófica que subyace al programa de naturalización de las emociones, a saber, la pretensión de que necesitamos una teoría naturalizada de la intencionalidad de las emociones.

Con ello es probable que todavía no salgamos del atolladero. Sin embargo, este trabajo es un pequeño paso en un proyecto todavía mayor, a saber, la elucidación del concepto de emoción. Lo que se busca no es dar una salida definitiva ni una terapia con la que podamos abandonar toda teoría naturalista de las emociones. Por lo pronto basta con sacar a la luz algunas confusiones conceptuales y entrever brevemente su origen. Así, tendremos presentes al menos algunas de las dificultades que atraviesan a estas teorías, y así tener más claridad sobre nuestras empresas científicas y filosóficas alrededor de nuestras emociones.

1. El programa de naturalización de las emociones

En este capítulo agruparé las teorías naturalistas de la emoción de Descartes, James y Prinz bajo la noción de "programa de investigación científica". La construcción de tal categoría se ajustará a la metodología propuesta por Lakatos para la reconstrucción racional de algunos episodios de la historia de la ciencia. En consecuencia, propondré un núcleo firme que atraviese las tres teorías y mostraré uno de los movimientos del cinturón protector que resultará importante para el desarrollo de la naturalización de la intencionalidad que trataré en el capítulo siguiente. Ver las tres teorías como un solo programa constituye el primer paso en mi argumentación, pues permite juntarlas bajo una misma categoría y permite evaluar con mayor claridad cómo las críticas de Wittgenstein se aplicarían al corazón del programa en su conjunto.

1.1. La metodología de los programas de investigación científica

La metodología de Lakatos. La metodología de los programas de investigación científica propuesta por Lakatos (1978/2010a, 1978/2010b) ofrece una herramienta interesante de reconstrucción racional de la historia de la ciencia.

Esta metodología historiográfica combina elementos del falsacionismo metodológico de Popper y del convencionalismo de Kuhn, siendo una alternativa intermedia entre ambas propuestas en filosofía de la ciencia. De allí que Lakatos llame a su posición *falsacionismo metodológico sofisticado*: falsacionismo, al aceptar las críticas de Popper al verificacionismo, y sofisticado, pues evita las ingenuidades que le atribuye Kuhn a Popper.[3]

Para Lakatos, lo que ha de evaluarse en la reconstrucción de la historia de la ciencia no es una teoría, sino una serie de teorías con cierta continuidad. A esta serie continua de teorías Lakatos las llama un "programa de investigación". Así las cosas, la metodología propuesta buscaría principalmente sacar a la luz la continuidad entre distintas teorías. Esta continuidad es capturada en la metodología bajo la categoría del "núcleo firme" del programa de investigación. Esta categoría recoge el conjunto de presuposiciones metodológicas y metafísicas, definiciones y demás ideas fundamentales del programa. El núcleo firme guía la investigación que se hace en el marco del programa, señalando los problemas, preguntas y métodos de solución válidos para el programa (lo que Lakatos llama la "heurística positiva").[4]

Ahora bien, como puede notarse, el núcleo firme es el corazón del programa. Esto implica, según Lakatos, que él

---

[3] Para la discusión sobre estas dos propuestas y la defensa de la metodología de Lakatos, véase Lakatos (1978/2010a).

[4] La adopción de un núcleo firme es uno de los elementos que se deciden por convención.

no es sometido nunca a escrutinio por parte de los científicos; al contrario, el núcleo firme ha de defenderse de toda evidencia y descubrimiento que pueda haber en su contra.[5] Esto garantiza que el programa de investigación acomode las anomalías y amplíe el espectro de las explicaciones que ofrece. Para lograr esta acomodación, un programa de investigación cuenta con un conjunto de hipótesis auxiliares que pueden adoptarse, modificarse o eliminarse con el fin de proteger el núcleo firme. A estas hipótesis Lakatos las llama "cinturón protector". Ellas son aquellas que el científico podrá falsear mediante experimentos y rechazar o moldear en caso de que una anomalía lo exija, a condición de mantener el núcleo firme incólume ante las instancias falsadoras.

La metodología de los programas de investigación permite también ver los cambios entre las teorías en el interior de un programa como cambios progresivos o regresivos. Cuando una teoría explica más hechos que su predecesora, esto es, cuando contiene bajo sus expectativas un exceso de hechos (exceso empírico), hay progreso teórico; cuando este exceso empírico es confirmado, hay progreso empírico. La conjunción de progreso teórico y progreso empírico es lo que Lakatos llama un "cambio de problemática progresivo". Cuando, por el contrario, un programa ya no explica

---

[5] Más propiamente, nunca habrá evidencia que se interprete como contraria al núcleo firme, toda vez que el núcleo firme ofrece las claves de interpretación de la experiencia. Uso "evidencia en contra" aquí como una experiencia inesperada de la que el programa, a la hora de encontrarla, no ofrece explicación. No obstante, el éxito del programa reside en acomodar esta experiencia ajustando sus hipótesis auxiliares.

hechos nuevos, cuando sus explicaciones solo son posteriores a los hechos y no hay ya capacidad de predicción, hay un cambio regresivo.

En síntesis, la metodología de los programas de investigación científica nos invita a ver la historia de la ciencia como la historia de teorías con una continuidad metodológica que llamamos "núcleo firme". Este núcleo contiene aquellas bases del programa que no es sometido nunca a falsación, y que delimita los problemas y preguntas de las que se ocupará la investigación científica ("heurística positiva"). Para salvar el núcleo firme se hacen modificaciones al conjunto de hipótesis auxiliares, al "cinturón protector". Cuando estos movimientos logran anticipar hechos nuevos, acomodar anomalías y predecir novedades, hay progreso teórico; cuando esto se confirma, hay progreso empírico. Finalmente, cuando hay predicción y confirmación de hechos novedosos, cuando hay progreso teórico y empírico, hay un "cambio de problemática progresivo".

Entre las ventajas de usar esta metodología, además de permitir ver progreso en la historia de la ciencia, Lakatos señala la inclusión de tesis metafísicas como parte de la ciencia y la posibilidad de incluir en la racionalidad científica la defensa constante de ciertas tesis incluso a pesar de las anomalías (1978/2010a, p. 59). Para nuestro caso, la metodología de los programas de investigación científica permite sacar a la luz la continuidad que hay entre Descartes, James y Prinz, y ver en las diferencias entre ellos el progreso de un programa. Dado que la metodología permite incluir en el trabajo científico tesis metafísicas, ella tiene la virtud

de permitirnos incluir en un programa de investigación la discusión filosófica que existe entre estos tres autores, incluso si ella no es directamente una investigación empírica. Además, la metodología nos permitirá ver en el problema de la intencionalidad uno de los problemas que debe atacarse dentro de este programa. Con esto en mente, pasemos a la reconstrucción de estas teorías y la construcción del "programa de naturalización de las emociones".

1.2. La teoría de las pasiones de Descartes

En muchos sentidos, Descartes es considerado el padre de la modernidad. Sus investigaciones en filosofía influenciaron más de un siglo de pensamiento, lo que lo hace uno de los filósofos más importantes de la historia. No obstante sus grandes logros en filosofía, Descartes también fue un reconocido hombre de ciencia. Más aún, su obra filosófica cubre una parte pasajera de su obra completa. En una carta a la princesa Isabel de Bohemia reconoce que las querellas metafísicas solo han de ocupar una parte de la vida de una persona. Para él, si bien es importante dedicarse en algún momento a la metafísica y la filosofía, es vital dedicarse también a lo que nos otorgan nuestros sentidos y el mundo (AT III 695).[6] En resumen, además de hacer filosofía, hay que hacer ciencia.

La teoría de las emociones de Descartes, extraída de su texto *Las pasiones del alma* (1649), hace parte de este

---

[6] En adelante citaré la obra de Descartes siguiendo la paginación de Adam y Tannery (AT). Véase Descartes (1887/1913).

programa científico. Allí Descartes nos ofrece una teoría mecánica sobre el funcionamiento de las emociones. Esto significa, entre otras cosas, una teoría sobre cómo el cuerpo provoca pasiones y emociones en el alma, cuál es la función de estas y cómo podemos identificarlas como parte del mundo natural. De ahí que para aproximarnos a la postura cartesiana sobre las emociones, debemos aclarar primero cómo concibe el autor el cuerpo humano y su lugar en la naturaleza, su anatomía, fisiología y cómo esto se aplica al caso de nuestro interés.

El cuerpo como máquina. Lo primero que diremos sobre la ciencia de Descartes es que el autor concibe el mundo natural como una gran máquina. El mundo cuenta con leyes que han sido impuestas por Dios desde su creación, y que permanecen inmutables como él. Este es uno de los pilares de lo que llama *la fábula del mundo* (AT X, 37).[7] Esta fábula cuenta, primero, con el desarrollo de una física y una cosmología que explica la naturaleza del espacio, la creación y evolución del universo y las leyes que gobiernan el mundo.

[7] No es claro todavía si esta fábula tiene pretensiones de ser una teoría sobre cómo es el mundo en sí mismo, si es un movimiento retórico para evitar presiones políticas o si en realidad es solo una construcción ficticia de un mundo que Dios pudo haber creado pero que no necesariamente creó. Descartes comenzó a escribir *El mundo*, compuesto por el *Tratado de la luz* y el *Tratado del hombre*, hacia 1629. Sin embargo, su primera publicación fue póstuma. El primero de los tratados fue publicado en 1664 (casi cuarenta años después de comenzado) y el segundo, dos años más tarde. Hay quienes sugieren que esta obra se retiró de la imprenta o fue abandonada tras la condena de Galileo en 1633 (Gaukroger, 1998/2004, p. viii). En cualquier caso, este mundo y esta fábula, tras su desarrollo, resultan pareciéndose a nuestro mundo natural, de modo que esta funge aquí como una teoría general del universo.

Luego Descartes pasa a ocuparse de la manera cómo funciona el hombre, esto es, cómo funciona el cuerpo humano, cuál es su naturaleza y cómo este encaja en el mundo físico. La primera parte está desarrollada en el *Tratado de la luz* (1664), la segunda, en el *Tratado del hombre* (1666).[8]

Ahora bien, hemos dicho que en esta fábula el universo es una máquina que cuenta con leyes universales e inmutables, impuestas por Dios desde su creación. Cabe aclarar que la noción de 'máquina' que usa Descartes puede diferir un poco de la nuestra actualmente. Cuando pensamos en máquinas, muchas veces pensamos en computadores o en aparatos que usan corrientes eléctricas. Sin embargo, Descartes, lejos de pensar en mecanismos eléctricos, piensa en mecanismos hidráulicos. Así, la metáfora del cuerpo y el universo como máquina debe pensarse en términos de un gran mecanismo que funciona a partir del movimiento de fluidos. De ahí que en el universo cartesiano, toda explicación del mundo como máquina, esto es, toda explicación *mecanicista*, debe ser reducible a las nociones de 'extensión' y 'cantidad de movimiento'. No entraré de lleno en los detalles de este tipo de explicación ni de la cosmología cartesiana. Por ahora basta con advertir que la metáfora con la que Descartes arma su ciencia tiene más parecidos con una serie de movimientos de fluido que con los computadores en los que podríamos pensar hoy en día.

---

[8] Se cita de la versión de Adam y Tannery, pero además se incluye en la bibliografía la edición en inglés que incluye ambos tratados. Véase Descartes (1998/2004).

Ahora bien, el cuerpo humano, en tanto físico, hace parte de esta gran máquina hidráulica. Por lo tanto, el cuerpo también debe tener leyes sobre su funcionamiento. En otras palabras, si el mundo es una máquina, el cuerpo humano es también una máquina que participa de esta máquina mayor. De ahí que se hable de una explicación mecanicista del cuerpo. En esta línea se desarrolla la teoría de las pasiones como una teoría en la que las pasiones han de encajar como función de la máquina que es el cuerpo. Hay que anotar, preliminarmente, que las pasiones incluyen, aunque no de manera exclusiva, a las emociones, de modo que hemos de avanzar primero en el campo de las primeras.

La anatomía del cuerpo humano. En la primera parte de LPA, Descartes resume lo que ya ha presentado antes en el *Tratado del hombre*. Primero, Descartes le atribuye al corazón la función de calentar la sangre. Al calentar la sangre, esta se expande y expande consigo el corazón. Esto genera un bombeo, de modo que la sangre sale disparada hacia todas las partes del cuerpo. Aquellas partículas que van con mayor velocidad van a parar al cerebro, en donde pasan por pequeños poros que las filtran. Al ser filtradas, quedan partículas de sangre muy pequeñas con una gran velocidad a las que Descartes llama *espíritus animales.*

Estos espíritus animales viajan desde el cerebro a través de los nervios para mover las distintas partes del cuerpo (inflando unos músculos y desinflando otros, por ejemplo) y para reportar todo tipo de información sensorial. Para Descartes, los nervios se entienden a la manera de pequeños

tubos que permiten justamente el flujo de espíritus animales. En su interior hay un filamento muy delgado que conecta los órganos sensoriales con una región en el cerebro, a saber, la glándula pineal —o, para mayor precisión, sus vecindades—. Los espíritus animales lubrican el paso por el que este filamento transmite información de manera instantánea a la glándula pineal, moviéndola e inclinándola de un lado para el otro.

En el caso de la percepción visual, por ejemplo, los rayos de luz que van desde el objeto hasta el ojo llegan a la retina, que conecta con el nervio óptico. En el interior del nervio óptico están estos filamentos que, dependiendo de la intensidad de la luz, el color, y otras propiedades, se mueven de determinada manera (se tensionan y rotan de una u otra forma). Al moverse, los filamentos abren y cierran algunos poros en el cerebro a los que llegan los nervios, y permiten un flujo mayor —o menor— de espíritus animales. Estos pasan por dichos poros y mueven la glándula pineal en una dirección particular. Este movimiento es entonces percibido por el alma como la imagen del objeto.

Esta es, de manera resumida y cortada, la forma en que Descartes explica las funciones del cuerpo. El caso de las emociones será muy similar. Para el autor, las pasiones (entre las que están las emociones) también son movimientos de los espíritus animales que causan movimientos de la glándula pineal que luego el alma percibe como alegría, tristeza, etc. Así, Descartes intentará acomodar las emociones al funcionamiento mecánico del cuerpo, de modo

que pueda ofrecernos una explicación de estas como parte del mundo natural.[9]

La clasificación de las pasiones. Descartes comienza su tratado distinguiendo entre las pasiones y las acciones. Las pasiones son aquellos estados en los que el alma recibe información pasivamente; por el contrario, las acciones son aquellos eventos que un sujeto hace que sucedan, o en los que el alma envía información al cuerpo y genera movimiento. Por ejemplo, la sensación de calor que siento al estar cerca al fuego, dado que es información que mi alma recibe de manera pasiva, es una pasión. En cambio, el movimiento de mi brazo, en tanto que es causado e iniciado por mí, es una acción. Es importante anotar que siempre que hay una pasión, hay un agente activo que la genera —y, de la misma manera, siempre que hay una acción, hay un agente pasivo sobre el que recae la acción—; en otras palabras, una pasión siempre implica una acción, y viceversa.

Ahora bien, entre las pasiones encontramos tres tipos diferentes. Algunas son referidas a objetos externos, como la percepción de un árbol frente a mí; otras son referidas

---

[9] Como en otros aspectos de la ciencia y la filosofía cartesiana, este tipo de estudio marca una diferencia importante con la aproximación aristotélica. En el caso de Aristóteles, las emociones aparecen como cosas a través de las cuales cambiamos nuestros juicios (n.d./2005, 1378a). El estudio que se hace de las emociones es, entonces, un estudio cognitivista, en donde se evalúan las razones por las cuales una persona siente una emoción y cómo estas llevan a cambiar los juicios. No hay, en este sentido, una sistematización de las emociones, ni un estudio natural de las mismas. Es en esto en lo que Descartes marca la diferencia y se convierte en el iniciador de un programa de investigación en donde las emociones se estudien desde la ciencia natural.

a estados internos del cuerpo, como el dolor que siento cuando una piedra cae sobre mi pie. Las terceras son aquellas que Descartes califica como "referidas al alma". Estas últimas son las emociones. Ellas son pasiones, por un lado, en tanto que el alma las siente con independencia de su voluntad, es decir, es pasiva frente a ellas y, por otro lado, son mantenidas y fortalecidas por los movimientos de los espíritus animales. Además, ellas son referidas al alma, en tanto que no tienen por objeto algo externo o interno al cuerpo. Por ejemplo, no me siento triste por la mesa (objeto externo), ni siento ira por mi pie (sensación interna); mis emociones refieren a eventos como la muerte de mi gato o la recepción de un regalo. En este sentido, las emociones refieren al alma para distinguirlas de las otras dos clases de pasiones (LPA, a. XXVIII).[10]

La función biológica de las emociones. Después de distinguir entre pasiones y acciones y de agrupar a las emociones como aquellas pasiones que refieren al alma, Descartes procede a describir más detenidamente en qué consisten las emociones, cuál es su función y cuál su clasificación. Como hemos dicho, las emociones para Descartes son pasiones, i.e., eventos percibidos por el alma, pero que no refieren a un objeto externo o interno. En tanto que pasiones, las emociones deben ser causadas y mantenidas por los movimientos de los espíritus animales, de modo que cada

---

[10] Descartes retoma aquí una nota etimológica sobre la palabra 'emoción', que remite a *movimiento*. Las pasiones que refieren al alma pueden llevar el nombre de emociones, toda vez que tienen el poder de movernos a una acción determinada (e.g., el miedo tiene el poder de llevarnos a correr para la huida).

una de las emociones tendrá una descripción en términos fisiológicos.

Esta descripción, recordemos, debe darse en términos mecanicistas, esto es, hace uso de la metáfora del cuerpo como una máquina (hidráulica, en el caso de Descartes). De ahí que la descripción de las emociones en términos fisiológicos debe incluir una función que estas han de cumplir para el funcionamiento de dicha máquina. En otras palabras, la explicación de las emociones debe decirnos para qué están ellas en el cuerpo, para qué Dios las ha instituido allí. La respuesta de Descartes a esta pregunta es central para el programa naturalista, pues le atribuye a las emociones una función relacionada con la supervivencia. Luego de aclararnos que su tratado se referirá a las emociones, en tanto que pasiones del alma, Descartes afirma:

> [...] es necesario advertir que el principal efecto de todas las pasiones [emociones] en los hombres es que incitan y predisponen su alma a querer las cosas para las que preparan su cuerpo: de suerte que el sentimiento del miedo incita a querer huir, el del arrojo a querer combatir, y lo mismo los demás (LPA, a. XL, AT 359).

Esto anticipa lo que luego James atribuirá a las emociones desde la teoría de la evolución de Darwin: la idea de que las emociones cumplen la función de preparar nuestro cuerpo con presteza para su supervivencia. El miedo, por ejemplo, sirve al cuerpo para que pueda huir rápidamente del peligro; el amor y el deseo, para llevarnos a lo que es bueno

para nosotros; la admiración, para guardar en la memoria aquello que es importante conservar, etc. Es importante anotar, sin embargo, que Descartes todavía no cuenta con la teoría de Darwin, de modo que la atribución de funciones debe recaer, no en la evolución, sino en el plan divino. En otras palabras, las emociones no tienen funciones para una ventaja evolutiva, sino porque Dios lo ha instituido así (LPA, a. XXXVI, AT 357).

Descartes nos ofrece un ejemplo para ilustrar lo dicho hasta ahora. Supongamos que nos encontramos con una criatura de forma espantosa, frente a la cual sentimos temor. La explicación de por qué sentimos temor va desde que percibimos la imagen de la criatura —desde que se forma una imagen[11] de ella en la glándula pineal— hasta que reconocemos esta emoción como temor. Recordemos, primero, que la imagen de esta criatura se forma gracias al movimiento de los espíritus animales sobre esta glándula. De ahí, dice Descartes:

> [Los espíritus animales] se trasladan, en parte, a los nervios que sirven para volver la espalda y mover las piernas para huir, y, en parte, a los que ensanchan o estrechan de tal manera los orificios del corazón, o bien que agitan de tal modo las otras partes de donde se le envía la sangre, que esta sangre, rarificándose de forma distinta

[11] Este es el término usado por Descartes para la representación que se forma después de que los rayos de luz entren al ojo y causen movimientos en el nervio óptico. Véase LPA, a. XXXVI, AT 357.

> a la acostumbrada, envía al cerebro espíritus apropiados para mantener y fortalecer la pasión del miedo [...]. Pues el solo hecho de que los espíritus entren en esos poros provoca un movimiento particular en esa glándula, el cual ha sido instituido por la naturaleza para hacer sentir al alma esa pasión (LPA, a. XXXVI, AT 357).

La descripción que tenemos aquí del miedo, entonces, es una descripción en términos del movimiento de la sangre y de la glándula pineal para generar una respuesta inmediata frente a un espanto. Así, vemos la forma en la cual las emociones se describen en términos fisiológicos, como el movimiento de la sangre a través del cuerpo para generar una reacción y, además, cómo estas suponen una reacción para la supervivencia.

Ahora bien, si las emociones responden a movimientos preestablecidos en la glándula pineal, ¿cómo es que personas diferentes pueden reaccionar de maneras distintas? Por ejemplo, ¿cómo puede una persona sentir miedo a las alturas y otra no? En principio, la percepción es exactamente la misma. Si seguimos el razonamiento anterior, deberíamos concluir que los movimientos en sus glándulas pineales sería el mismo, de modo que sus cuerpos deberían reaccionar igual. No obstante, Descartes aclara que las emociones pueden controlarse y dominarse mediante el hábito. De manera general, la idea radica en que a través del hábito cambiamos nuestra disposición cerebral para permitir o impedir el paso de los espíritus animales en determinadas circunstancias. A lo largo de nuestra vida puede que haya poros en el cerebro

que se abran con mayor o menor facilidad, dependiendo de nuestras experiencias y pensamientos anteriores. Así, una persona puede acostumbrarse a las alturas de modo que su reacción cerebral sea diferente a la de una persona que no está acostumbrada (LPA, a. L, AT 368).[12]

En síntesis, la teoría cartesiana inaugura una manera de pensar las emociones en términos naturalizados, esto es, en términos de reacciones fisiológicas y cerebrales. Esta primera teoría supone que las emociones son percepciones referidas al alma, esto es, que no son referidas a objetos externos o a partes particulares del cuerpo (como el dolor). Estas, en tanto que son una clase de pasión, no dependen de la voluntad del alma, sino de la disposición y el movimiento del cuerpo. En particular, para Descartes las emociones dependen del movimiento de los espíritus animales. Tienen la función de preparar al cuerpo para su bienestar y supervivencia, de modo que están predeterminadas —o instituidas— por naturaleza (divina) para protegernos del peligro o llevarnos a buscar lo que es bueno para nosotros. Sin embargo, esta predeterminación puede ser reconfigurada mediante el hábito, de modo que distintos sujetos puedan experimentar diferentes reacciones emocionales.

Más adelante, tras la aparición de la teoría de la evolución de Darwin, la teoría tomó un rumbo que permitió el progreso del programa. En lugar de atribuirle la función de las emociones a un plan divino, ya era posible hacer recaer

[12] Para un análisis sobre cómo es posible que el hábito cambie nuestras reacciones emocionales en la teoría cartesiana, véase Shapiro (2003).

funciones naturales en la evolución de las especies. Este avance, propuesto principalmente por James, permitió incluir las explicaciones de la teoría darwinista en una teoría naturalizada de las emociones.

1.3. La teoría James-Lange

El programa de investigación naturalista tuvo su mayor impulso tras los aportes de William James y Carl Lange, a quienes se les atribuye la formulación simultánea de la definición de las emociones como la sensación de reacciones corporales. De ahí que el programa de investigación que estamos tratando se conozca generalmente como la teoría James-Lange. Para los motivos de este texto nos centraremos en la versión de James, dado que es a él a quien Wittgenstein dirigirá sus anotaciones. No obstante, es importante reconocer el papel de Lange en el avance de esta teoría, así como de las reservas de James con algunas de sus ideas.[13]

Como hemos señalado, la idea central de la teoría James-Lange es la identificación de las emociones con la sensación de cambios en el cuerpo. La definición clásica de James versa de la siguiente manera: "[…] los cambios corporales siguen directamente a la percepción del hecho excitante, y […] nuestra sensación de los mismos cambios conforme ocurren ES la emoción" (PP, p. 915, énfasis en el original;

[13] James pensaba que Lange daba demasiada importancia a las reacciones viscerales en la emoción. Para el primero, podía haber otros tipos de reacciones corporales, incluyendo reacciones netamente cerebrales. (1884, nota al pie 2, pp. 199-200).

1884, p. 189).[14] El ejemplo paradigmático del autor es el temor que siente una persona frente a un oso. Supongamos que estoy en el bosque y veo que un oso se acerca hacia mí. En ese momento percibo al oso y mi cuerpo reacciona de inmediato. Siento agitación, adrenalina, probablemente mi cara se torne pálida y mis pupilas se dilaten. Para James, todas estas reacciones ocurren justo tras la percepción del oso. En este punto soy consciente de todas esas reacciones y esta consciencia, este *sentir* esas sensaciones, es lo que llamo *temor*.

Esta definición lleva a James a pensar que el orden de los eventos en un episodio emocional es primero, la percepción de un hecho (en este caso, ver el oso); segundo, la reacción de mi cuerpo (elevación del ritmo cardiaco, secreción de adrenalina, etc.) y tercero y último, la identificación de esta respuesta corporal como temor.[15] De allí la insistencia de James en que:

> La hipótesis que vamos a defender dice que [...] la enunciación más racional es que sentimos tristeza porque lloramos, furia porque golpeamos, miedo porque

[14] Citaré los PP y, donde sea posible, el artículo de 1884. Los PP reproducen casi todo este artículo, de modo que en la mayoría de los casos el lector puede referirse a cualquiera de estos textos.

[15] James intenta atacar aquí la idea según la cual las reacciones corporales son producto de la emoción,y, por tanto, vienen después de esta. Bajo este orden, tras percibir el oso, lo juzgo como peligroso y siento temor; luego de esto mi cuerpo reacciona. Este es el orden defendido por varios teóricos e investigadores de las emociones, entre los cuales se destacan cognitivistas como Solomon (1973/1980).

> temblamos, y no que lloramos, golpeamos o temblamos porque estamos tristes, irritados o temerosos, según el caso (PP 915; 1884, 190).

Aquí, el llanto, el golpe o el temblor funcionan como aquellos cambios en el cuerpo que permiten identificar mi estado como un estado de tristeza, de furia o de miedo, respectivamente. De ahí que James formule esta hipótesis haciendo énfasis en que las reacciones corporales van primero que las emociones. En otras palabras, lo que quiere resaltar James es que sentimos una emoción porque tenemos una respuesta corporal, y no que esta última es generada por un estado mental distinto que llamamos emoción.

El protagonismo del cuerpo. Como puede notarse, James también le da un papel central al cuerpo y las reacciones fisiológicas que ocurren durante las emociones. Investigar las emociones desde la teoría James-Lange es estudiar la forma en que nuestro cuerpo reacciona mientras las sentimos, cómo estas reacciones se comunican con el cerebro para generar la sensación correspondiente y cuál es la fenomenología correspondiente a dicha sensación, es decir, cómo se vivencian estas reacciones. Por ello este programa marca un camino de investigación no solo en la psicología, sino también en la fisiología y, con mayor fuerza en la actualidad, en las neurociencias.[16]

---

[16] Como veremos más adelante, los refinamientos de Prinz al programa se valen en buena medida de los avances recientes en neurociencias.

Hemos dicho antes que James identifica a las emociones con la sensación de cambios en el cuerpo. En esto James parece ir por la misma vía de Descartes. Descartes realiza una explicación fisiológica de las emociones que invita a pensarlas como la percepción de reacciones corporales. Más aún, tanto las percepciones visuales como las emociones están clasificadas bajo las pasiones, de modo que los mecanismos son bastante similares. Sin embargo, la clasificación de las pasiones en tres clases lleva a Descartes a pensar en las emociones como pasiones que refieren al alma, por contraste a las pasiones que refieren a objetos externos y a las que remiten a estados internos en el cuerpo. En cambio, la teoría de James simplifica esta clasificación y subsume las emociones dentro de esta última categoría, i.e., las pasiones que refieren a estados internos.[17] Esto no solo genera una clasificación más simple, sino que evita la postulación de un alma a la cual hacer referir las emociones. En otras palabras, la teoría de James no requiere de un alma que sienta; solo requiere de un cuerpo capaz de tener y ser consciente de estas sensaciones.

La atribución de las emociones a sensaciones internas también puede notarse en la forma en la que James plantea

[17] La clasificación de las emociones como "pasiones del alma" puede dar lugar a confusiones sobre el papel de la cognición en estas. Por ejemplo, Gary Hatfield (2007a) sostiene que la distinción entre James y Descartes radica precisamente en que Descartes, al referir las pasiones al alma, requiere de la cognición para dar sentido a las emociones. En este orden de ideas, la teoría de Descartes no haría recaer las emociones enteramente en el cuerpo, como sí lo hace la de James.

el problema en 1884, y algunas de las secciones presentes en los PP. James abre su artículo preguntándose si existe un solo centro cerebral para las emociones, o si existen varios. Recordemos que para la época de James, nuestro entendimiento del cerebro es muy distinto al de Descartes. James ya cuenta con una teoría del cerebro que reconoce en los nervios, no tubitos llenos de espíritus animales, sino una suerte de cableado por el que pasan corrientes eléctricas. Estos nervios, a su vez, ya no conectan con una sola parte del cerebro (como la glándula pineal de Descartes), sino que conectan con diversas partes del cerebro; ello lleva a distinguir entre varias partes y funciones en este órgano. En particular, se distingue entre centros sensoriales y motores (PP, c. II). De ahí que James pregunte si hay un centro separado para las emociones, un centro emocional, si se quiere, o si podemos reducir las emociones a los centros ya disponibles. Su respuesta, finalmente, apunta a que es posible reducir las emociones a actividades de los centros sensoriales de modo que hagan parte de nuestras demás sensaciones internas.

Dos argumentos importantes a favor del cuerpo. James insiste constantemente en que no se puede abandonar el cuerpo al hablar de las emociones. Entre estas insistencias encontramos dos argumentos importantes. El primero se basa en imaginar un caso en el que una persona siente una emoción de la cual abstraigamos toda sensación corporal. Tal abstracción, para James, elimina toda emoción y nos deja con un simple juicio sobre una situación. En palabras de James:

> Si imaginamos una emoción fuerte, y luego intentamos abstraer de nuestra conciencia todas las sensaciones de sus síntomas corporales característicos, encontramos que no nos queda nada, ninguna "cosa-mental" de la cual la emoción pueda estar constituida, y un estado frío y neutro de percepción intelectual es todo lo que permanece (PP, p. 451; 1884, p. 193).

Este experimento mental marca lo que Prinz luego llamará la "hipótesis de la corporeidad" (*embodiment hypothesis*). Según esta hipótesis, no es posible entender las emociones sin cambios corporales. Para James no solo las emociones involucran cambios en el cuerpo, sino que es imposible imaginar siquiera un caso en el que esto no fuese así, al menos para los seres humanos.[18] Aquí, no se trata de una simple contingencia, sino de una condición necesaria de todo episodio de emoción. Pensemos, por ejemplo, en una tristeza profunda, y a continuación intentemos pensarla sin el nudo en la garganta, o en el temor sin la agitación del corazón. Es esto lo que, según James, es imposible hacer; estas emociones son precisamente estas reacciones.

Este primer argumento muestra que es impensable la emoción sin cambios corporales. No obstante podría pensarse en un caso en el que el paciente presente dichos cambios, pero no los sienta. Para James, un caso de estos podría presentarse con un paciente anestesiado que, si bien tendría

[18] Para James es posible pensar esto para otro tipo de organismos o entidades, pero insiste en que no es posible para seres humanos. Véase James (1884, p. 194).

las reacciones corporales correspondientes, no podría ser consciente de ellas. En un caso así es posible que el cuerpo siga segregando las sustancias químicas respectivas, moviendo los músculos como es de esperarse, etc., pero la anestesia haría que la persona no fuera consciente de ninguno de estos cambios. Si tuviésemos un caso de estos y el paciente todavía mostrara respuestas emocionales —dice James— tendríamos que abandonar la teoría (1884, p. 203). En este sentido, dicho caso se configura como una posible instancia de falsación del programa de investigación.

Además de esto, James afirma que es posible que haya casos en los que la emoción no tenga ningún objeto particular. Un caso de estos sería, por ejemplo, uno en el cual la persona siente temor, pero no de algo en particular; simplemente lo siente. Estos casos, para el autor, son casos patológicos, pero en los cuales hay, sin embargo, una respuesta emocional (1884, p. 199-201). La importancia de estos casos radica en que no solo es la emoción una sensación de cambios en el cuerpo, sino que es únicamente esta sensación la que constituye una emoción. Incluso si nada más hay sensación y no existe ningún tipo de objeto, nada que temer o nada por lo que estar triste, debemos llamar a esto emoción si hemos de seguir a James. Este punto será de vital importancia más adelante, cuando discutamos aspectos relacionados con la intencionalidad de las emociones en la teoría de James.

La percepción del hecho excitante. Hemos hablado de la primacía que tiene el cuerpo en esta teoría, pero hace falta entender cómo se generan las respuestas en él. En particular, es importante aclarar cómo se relaciona la percepción de

determinados eventos con nuestras respuestas corporales. Tomando de nuevo el ejemplo de James, ¿qué hay de particular en la percepción del oso que hace que mi cuerpo reaccione inmediatamente para que yo sienta temor?

En Descartes la explicación se le atribuía a ciertas funciones naturales que Dios había instituido en el cuerpo humano. Por ejemplo, podríamos decir que Dios ha preparado nuestro cuerpo (especialmente nuestro cerebro) para reacciones frente a la percepción del oso. Sin embargo, esto hace necesaria la inclusión de Dios y un diseño inteligente de las emociones. En ello James supone un progreso importante del programa. Para él, buena parte de nuestras respuestas emocionales son disparadas como respuesta evolutiva en sentido darwinista, es decir, como una respuesta natural a nuestro ambiente. James tiene en alta estima la teoría de la evolución de Darwin (Palencik, 2007, p. 773) y la usa para explicar cómo podemos dar cuenta de reacciones inmediatas como la del oso. El temor que sentimos tras la percepción de un oso, para el autor, sería una respuesta para reaccionar rápidamente ante el peligro que representa este animal, una respuesta que representa una ventaja selectiva. De la misma manera piensa James de otras respuestas emocionales, como el amor de una madre hacia su hijo o el miedo a las alturas (1884, p. 191). Todas estas son formas en que nuestro cuerpo ha evolucionado para protegernos del peligro y asegurar nuestra supervivencia, de modo similar a un instinto.[19]

---

[19] El autor diferencia entre instintos y emociones al comienzo del capítulo sobre las emociones en los PP, mas no en el artículo de 1884. La distinción

Esta explicación da cuenta de casos en los que hay una necesidad de supervivencia y perpetuación de la especie. Sin embargo, hay muchos casos de respuestas emocionales en las que no es tan claro que medie la evolución, como sentir ira hacia un amigo porque me ha engañado, o sentir temor a estar desempleado. Estos casos requieren un tratamiento más cuidadoso que intentaré ofrecer en el siguiente capítulo. Por lo pronto, podemos adelantar lo siguiente: En el artículo de 1884 James sostiene que la asociación juega un papel importante en estos casos, pues sentimos estas emociones porque aprendemos a asociar determinados comportamientos con ciertas predisposiciones evolutivas. Así, el miedo al desempleo podría ocurrir porque asociamos la idea de estar desempleados con, quizás, la inseguridad; de modo similar, mi ira por el engaño de mi amigo puede ocurrir por una asociación entre el engaño y cierto daño que me fue infringido. Estas asociaciones son las que, en principio, permiten conectar las emociones con casos de nuestra vida social e intelectual (James, 1884, p. 195-196). En este sentido, James repite la explicación cartesiana que le atribuye estos casos al hábito, pero la complementa remitiendo a funciones que hemos adquirido por evolución.

A pesar de esta explicación, varios han considerado que James todavía se queda corto al dar cuenta de la intencio-

radica en que las emociones permanecen en la vida subjetiva de las personas, esto es, yo soy el único que siente mis emociones. En cambio, los instintos sí pueden entrar en relaciones externas al sujeto. Por ejemplo, un reflejo como patear ante un golpe en la rodilla supone un evento en el mundo (mi patada) y no una experiencia netamente interna. Véase PP (cap. XXII).

nalidad de las emociones. En principio, la aceleración de mi corazón y la dilatación de mis pupilas no son "sobre un oso" cuando siento temor por él. Los sentimientos y las reacciones corporales, de suyo, no dan cuenta de la intencionalidad. En ese sentido, podría necesitarse algo más que la simple descripción fisiológica para explicar por qué una emoción adquiere un objeto, por qué el temor es 'sobre el oso'. En otras palabras, si las emociones son reacciones corporales, pero estas por sí solas no son intencionales, hace falta explicar todavía cómo surge este rasgo.

Salvar la teoría de James de este problema no es tarea sencilla. Algunos teóricos han preferido optar por juntar la teoría de James con teorías cognitivistas, y mostrar que las emociones son sensaciones de reacciones corporales acompañadas de cierta clase de juicios.[20] Esto resulta, por supuesto, en la alteración del núcleo firme de la teoría, toda vez que cambiamos la definición central de las emociones como solo la sensación de reacciones corporales, agregando otros elementos. No obstante, hay quienes defienden todavía la teoría sin alterar esta definición. En particular, Prinz, en una muestra de tenacidad científica,[21] ha intentado enmendar este bache en la teoría de James con una teoría

[20] Para ejemplos de esto véase Damasio (1994) y Helm (2009).

[21] Bajo la metodología de Lakatos se puede explicar la actitud de los científicos para proteger el programa en el que trabajan. Incluso a la luz de problemas y dificultades duraderas, los científicos protegen el núcleo y prueban movimientos del cinturón protector mientras la teoría progrese. Esta tenacidad es la que hacemos recaer ahora en Prinz, toda vez que optó por apegarse al núcleo en lugar de abandonar el programa. Sobre la tenacidad de las teorías en los programas de investigación, véase Lakatos (1978/2010a, p. 116).

de la representación mental que permita dar cuenta de la relación entre la intencionalidad y las respuestas fisiológicas.

1.4. La teoría neo-jamesiana de Jesse Prinz

Entre las teorías contemporáneas de la emoción, una de las que más ha recibido atención ha sido la teoría naturalista de Jesse Prinz (2004b). Prinz busca rescatar la teoría James-Lange, identificando como problema crucial la naturalización de la intencionalidad (Prinz, 2004a). Este filósofo defiende la idea de que las emociones son la sensación de reacciones corporales (tesis básica de la teoría James-Lange), a la vez que explica cómo aspectos intencionales como creencias y deseos entran a jugar un papel en nuestras respuestas emocionales. Para ello, Prinz se vale de la teoría de la representación mental de Dretske, de modo que las emociones aparecen ahora como la representación de una relación de un objeto con el ambiente del organismo en virtud de sus estados corporales. Esto quedará más claro en lo que sigue. Por lo pronto, debo advertir que me centraré solo en algunos aspectos de la teoría de Prinz que irán al servicio de la discusión presente. Sin embargo, reconozco que esta teoría es de una gran complejidad y de ninguna manera queda agotada en la breve descripción que daré a continuación.

Una breve introducción a la psicosemántica. Hemos dicho que la teoría de Prinz se basa en la teoría James-Lange. La apuesta básica de esta última es que las emociones son la sensación de los cambios corporales mientras ellos ocurren —siguiendo la formulación de James (1884; 1890/1950)—.

Sentir ira es, en este orden de ideas, la sensación de una cantidad de cambios corporales de toda índole, desde aumento del ritmo cardiaco, respiración acelerada, hasta el flujo de adrenalina por el cuerpo. Para James no existe un estado mental separado que llamemos emoción; el estado de sentir estos cambios es lo que llamamos emoción. Sin embargo, esta teoría deja sin explicar el papel de elementos intencionales como creencias y deseos en nuestras respuestas emocionales. Por ejemplo, no se explica cómo una creencia como "Mi perro ha muerto" puede tener incidencia sobre mis estados corporales, de modo que yo perciba estos cambios y a esto lo pueda llamar "sentir tristeza". La relación entre lo corporal y lo intencional queda por fuera, de modo que es esto lo que Prinz intentará enmendar para fortalecer la teoría.[22]

Para ello, el autor se vale de la teoría de la representación mental de Dretske y así sostener que las emociones, siendo sensaciones de reacciones corporales, funcionan como representaciones de temas relacionales básicos (*core-relational themes*). Primero, aclaremos qué significa una representación mental en la teoría de Dretske, tal y como la presenta Prinz: De modo general, un sistema representacional es uno que tiene la función[23] de llevar información sobre alguna

---

[22] En otro artículo, Prinz ha defendido la teoría James-Lange de diversas objeciones y ha sugerido que este problema, el de la naturalización de la intencionalidad, es el que merece atención para corregir la teoría. Ver Prinz (2004a).

[23] Es importante señalar que un sistema representacional debe tener una función, pues de lo contrario solo sería un indicador. Si pensamos en los anillos de un árbol, ellos llevan información sobre la vida del árbol, pero no tienen esta función, por lo que solamente *indican* cuán viejo es el árbol, mas no lo representan. Lo mismo ocurre con el humo que sale de una fogata, que indica

propiedad (Dretske, 1995, p. 2). A su vez, una representación mental es un estado en un sistema representacional que se caracteriza por dos condiciones: *(1)* lleva información de aquello con lo que coocurre y *(2)* puede ser erróneamente aplicada. La coocurrencia de la que hablamos aquí es, en muchos casos, causal; una representación mental lleva información de aquello que la causó.[24]

Pensemos primero en el caso de un sistema representacional en general y luego aclaremos cómo esto aplica a nuestros estados mentales. Tomemos uno de los ejemplos clásicos de la literatura: un termómetro. Un termómetro es un sistema que tiene la función de llevar información sobre la temperatura, usando una barra de mercurio líquido que se expande o se contrae gracias a cambios en la temperatura del ambiente. Supongamos que cada vez que la temperatura del ambiente es 20°C, la barra mide diez centímetros —esto dependerá de la construcción particular del termómetro—, y que cuando la temperatura es 25°C, la barra se expande a quince centímetros. Cada longitud particular de la barra, en este caso 10 cm o 15 cm, será un estado del sistema. Estos estados son causados por la temperatura, es decir, la temperatura causa que la barra tenga

---

la presencia del fuego, mas no tiene la función de llevar información del fuego y por tanto tampoco califica como sistema representacional.

[24] En la versión de Dretske (1995), la intencionalidad queda explicada por la relación causal que hay entre el objeto y el sistema representacional. Esto quedará mejor tratado en el segundo capítulo. Cabe señalar que la noción de causalidad que expone Prinz no es una noción fuerte en el sentido de la conexión necesaria entre dos eventos, sino una idea de la causalidad como una probabilidad alta de que, dado un evento, se dé otro.

tal o cual longitud, por lo que decimos que la longitud de la barra *representa* la temperatura. Así, la longitud de la barra satisface la primera condición, la de llevar información de aquello con lo que coocurre. Ahora supongamos que nuestro termómetro sufre algún daño de modo que la barra se ve contaminada. La barra puede ahora expandirse y contraerse, representando algo de la temperatura, pero ahora representa de manera equivocada aquello que debe representar (ahora se expande a 15 cm cuando la temperatura es 20°C). En este caso, el sistema satisface la segunda condición: los estados del sistema pueden representar erróneamente. Con esto, tenemos que el termómetro califica como un sistema representacional, pues tiene la función de llevar información, sus estados llevan información de aquello con lo que coocurre y pueden representar de manera equivocada.

Ahora, Drestke quiere defender que podemos entender los estados mentales como estados de un sistema representacional, es decir, como estados que llevan información de aquello con lo que coocurren y que pueden ser aplicados con error. Un ejemplo de representación mental son los conceptos. Pensemos en el concepto "perro". Este concepto lleva información de las percepciones que han causado que tenga el concepto "perro", esto es, de todas las veces anteriores en las que he visto perros. Además, puedo aplicar el concepto equivocadamente para referirme a otros objetos pensando que son perros, por ejemplo, puedo ver un lobo en la lejanía y confundirlo con un perro. Aquí, la percepción del lobo causa, equivocadamente, una representación

mental de "perro". En este sentido, un concepto cumple con las dos condiciones de Dretske.

Además de estas dos condiciones, recordemos que un sistema representacional debe tener la función de llevar información y, por tanto, una representación mental debe tener una función. En palabras de Prinz, "una representación mental es un estado mental que ha sido *configurado* [*set-up*] para ser *disparado* [*set-off*] por algo" (2004b, p. 54). En este caso, el concepto "perro" ha sido configurado para dispararse frente a la presencia de perros, y puede ser erróneamente disparado, como en el caso del lobo. Esta función (e.g. la de representar los perros) puede ser fijada de dos maneras: por evolución o por aprendizaje. Cuando decimos que se configura una representación por evolución, decimos que hay estados mentales que generamos como respuestas evolutivas, como el reconocimiento de comida por parte de algunos animales (e.g. una rana reconoce en una mancha negra, una mosca, una fuente de comida). Por contraste, cuando decimos que se configura por aprendizaje, decimos que armamos representaciones por episodios de ostensión repetida en los que aprendemos que determinadas percepciones *representan* cosas. Este sería el caso del concepto "perro", que aprendemos a usar tras repetidos episodios de asociación entre distintas propiedades (hocico, patas, orejas…) y de agrupación de estas bajo un concepto.

Las emociones como representaciones mentales. Volvamos ahora al caso de las emociones. Prinz piensa que las emociones funcionan de manera similar a los conceptos, i.e., ellas también son representaciones mentales, lo que implica que

llevan información y pueden aplicarse erróneamente —además de que se configuran por evolución o por ostensión—. La primera pregunta aquí es: ¿de qué llevan información y a qué se pueden aplicar erróneamente? Prinz sostiene que las emociones representan —esto es, llevan información de o coocurren con— "temas relacionales básicos". Un tema relacional básico es una relación entre un organismo y su ambiente que atañe al bienestar de este organismo. Esta noción es tomada de Richard Lazarus, quien lista distintos temas relacionales básicos para distintas emociones. Por ejemplo, la ira es una relación de ofensa de algo hacia mí o hacia algo mío (Prinz, 2004b, 16).[25] En otras palabras, siento ira cuando me encuentro en una situación en la que hay una ofensa hacia algo mío o hacia mí (mi estado coocurre con la situación). Esta relación, como puede notarse en el ejemplo, versa sobre cómo algo puede afectarme negativa o positivamente, es decir, cómo algo puede afectar mi bienestar.

Tenemos entonces que una emoción lleva información (representa) una relación de mí como organismo con mi ambiente (un tema relacional básico). ¿Cómo funciona tal representación? Volvamos por un momento al caso del concepto "perro". Este concepto representa a los perros, pues se ha configurado para buscar ciertas propiedades de los perros, tales como un hocico, patas, etc. Esto es a lo que Prinz hace referencia cuando sostiene que el concepto *rastrea* estas propiedades; el concepto refiere a este conjunto de propiedades. No obstante, hay dos formas en las

---

[25] Para la lista completa ver Prinz (2004b), 16.

que un estado (mental) puede rastrear propiedades. Por un lado, podríamos decir que el concepto "perro" rastrea estas propiedades aparentes como el hocico y las patas, pero, además, podemos decir que rastrea otras propiedades, en este caso, no de la apariencia del perro, sino de su patrón genético (que es único para los perros) y por esto podemos distinguir a un perro de otra especie mirando solo su genética. A los primeros estados, Prinz los llama "detectores rastreadores de apariencia" (*appearance-tracking detectors*), pues rastrean propiedades de la apariencia de los perros y no propiedades esenciales; a los últimos los llama "detectores rastreadores de esencia" (*essence-tracking detectors*), toda vez que rastrean una propiedad esencial de los perros, como su genética.

Además de esto, Prinz distingue entre contenido «real» y contenido «nominal» de un estado. En el caso de los conceptos, su contenido real es el conjunto de propiedades esenciales del concepto; el nominal, el conjunto de propiedades (aparentes o esenciales) mediante las cuales identificamos lo que el concepto representa. Así, el contenido real del concepto "perro" es el conjunto de propiedades esenciales, como los genes o cualquier otra propiedad inherente a los perros; en cambio, el contenido nominal incluiría el hocico, las patas y demás propiedades aparentes de los perros. En este sentido, el concepto "perro" representa a los perros en virtud de su apariencia, independientemente de si conocemos o no las características esenciales de estos animales —aquí, el contenido real de los perros continúa siendo su conjunto de propiedades esenciales, pero el contenido nominal, el

contenido en virtud de lo cual identificamos a los perros, es su apariencia—. Así, el concepto "perro" sería un detector rastreador-de-apariencia.[26]

Las emociones también son detectores rastreadores-de-apariencia. Ellas representan temas relacionales básicos en virtud de ciertas apariencias que, si bien no son esenciales a estos temas, sirven para identificarlos. Estas apariencias son reacciones corporales. Así las cosas, las emociones son percepciones de reacciones corporales que, rastreándolas, identifican temas relacionales básicos. En otras palabras, los contenidos reales de la emoción son temas relacionales básicos, mientras que los contenidos nominales, mediante los cuales identificamos estos temas, son las reacciones corporales. Así como el concepto "perro" ocurre al tener la sensación de un perro, las emociones ocurren cuando tenemos una sensación corporal; a su vez, así como el concepto "perro" refiere a un conjunto de propiedades de los perros, las emociones refieren a temas relacionales básicos. Sentir ira, bajo esta teoría, es sentir una serie de reacciones en el cuerpo, reacciones que representan una relación con el ambiente (la de una ofensa hacia algo mío o hacia mí).

---

[26] Podríamos imaginar un caso en el que un grupo de biólogos ha vivido encerrado en un laboratorio recopilando información genética de distintas especies. Para ellos, el contenido nominal de los perros podrían ser los mismos genes, pues ellos identificarían los perros en virtud de estas propiedades esenciales. Es importante señalar, entonces, que el contenido nominal puede ser igual al contenido real, pese a que esto no siempre es el caso. De ser así, el concepto "perro" que tienen los biólogos sería un detector rastreador-de-esencia, pues el concepto representa no en virtud de apariencias, sino de la esencia de esta especie.

Cabe aclarar que, si bien los temas relacionales básicos son descritos mediante estructuras complejas, las emociones no tienen por qué tener esa misma estructura. La analogía que usa Prinz, tomada de Dretske, es la de los detectores de radar en los automóviles. Un detector de radar emite un sonido cuando hay un radar cerca, en este sentido, el sonido representa un radar. No obstante, el sonido no nos dice nada complejo acerca del radar, esto es, no lo describe; simplemente coocurre con él. De modo similar, las emociones no describen temas relacionales básicos, solo coocurren con estos temas. Estas son nada más la sensación de reacciones corporales (como el sonido del detector es solamente un sonido, y no es el radar mismo), pero representan una relación del organismo con su ambiente (como el sonido representa el radar).

Esta enmienda a la teoría James-Lange permite, en principio, solucionar el problema de la intencionalidad. Al introducir los temas relacionales básicos en las emociones, sin abandonar la tesis básica de que las emociones son la sensación de cambios corporales, se muestra cómo relacionar creencias y deseos con respuestas emocionales. Los temas relacionales básicos, como los trata Prinz, pueden construirse por aprendizaje o por evolución. Así, podemos aprender qué es peligroso y por tanto hacia qué sentimos temor, o podemos heredarlo evolutivamente.[27] Luego, si

[27] Entre los ejemplos de reacciones emocionales heredadas evolutivamente, Prinz menciona el temor a ciertos reptiles como las serpientes. El temor aquí es respuesta a un estímulo perceptual particular que estamos determinados a reconocer como peligroso. Ver Prinz (2004b), cap. 5.

podemos aprender ciertas asociaciones entre objetos con temas relacionales básicos, podemos entender cómo ciertas creencias pueden incidir en nuestras emociones. Puedo sentir ira porque mi amigo incumplió una promesa, toda vez que he aprendido a asociar el incumplimiento de promesas con algo ofensivo, y mi ira representa esta relación en la que algo me ofende. Puedo también sentir tristeza por perder un examen, pues también he aprendido a tomar esto como una pérdida, lo que corresponde con el tema relacional básico de la tristeza.

Este tratamiento de la intencionalidad que he presentado, sin embargo, es mucho más complejo en la teoría de Prinz. En el siguiente capítulo, en el que entraré de lleno en el tema de la intencionalidad de las emociones, trataré con mayor cuidado cómo la teoría de Prinz, así como las de Descartes y James, tratan de dar cuenta de dicha estructura intencional. Por lo pronto conviene cerrar el capítulo cumpliendo lo dicho al comienzo del mismo: construyendo un programa de investigación ajustado a la metodología de Lakatos. Esto será vital para el argumento que se intentará construir, toda vez que será lo que permitirá evaluar si es posible ampliar las críticas de Wittgenstein a James a todas las tres teorías en cuestión.

1.5. La construcción del programa de investigación naturalista

Hemos expuesto aquí tres teorías que comparten la misma aproximación a las emociones. Las etiquetas que pueden usarse para designar este tipo de teorías son variadas. Por

lo que hemos adelantado llamaré a estas teorías "naturalistas", en tanto que su principal objetivo es una descripción naturalizada de las emociones.

El núcleo firme. Como hemos advertido ya, el núcleo firme del programa naturalista sobre las emociones cuenta primero con la definición de las emociones como la sensación de cambios en el cuerpo. La formulación canónica de esta definición es la de James, que podemos citar nuevamente en gracia de la reconstrucción: "[…] los cambios corporales siguen directamente a la percepción del hecho excitante, y […] nuestra sensación de los mismos cambios conforme ocurren ES la emoción" (PP, p. 915; 1884, p. 189). En Descartes, las emociones aparecen como una subclase de las pasiones, que son cambios en el cuerpo frente a los cuales el alma es pasiva. Sin embargo, las emociones requieren una clasificación que las aparte de las sensaciones internas del cuerpo (como el dolor). James simplificará esta clasificación, y las reducirá a los mismos centros sensoriales mediante los cuales percibimos sensaciones en nuestro cuerpo. En otras palabras, para James no es necesario postular una tercera categoría de las pasiones, a la manera de Descartes. Para él, las emociones son ahora una subclase de las sensaciones internas. Sin embargo, ya en Prinz las emociones no son simples sensaciones, a la manera del dolor. Las emociones, siendo percepciones de cambios en el cuerpo, representan relaciones de bienestar, y nos dan información sobre aquello que las causa. En cualquiera de los tres casos, no obstante, la explicación sigue siendo en términos de reacciones en el cuerpo y algún modo de percepción. Prinz incluso resume

esta postura con la suya propia, de la siguiente manera: "Siguiendo la tradición de Descartes, James, Lange y, más recientemente, Damasio, creo que las emociones son estados internos que ocurren en respuesta a cambios corporales" (2004b, p. 242).[28]

El estudio de las emociones que implica este programa supone, por supuesto, entender el cuerpo humano, de modo que el programa naturalista tiene un especial énfasis en la fisiología. En el caso de Descartes, por ejemplo, la descripción se ha dado en términos de movimientos de los espíritus animales y la glándula pineal; el cuerpo aparece aquí como una máquina que se asemeja a una máquina hidráulica. Para James, las reacciones también se describen en términos de movimientos de la sangre, aunque ya no en términos de espíritus animales. Además, se introducen explicaciones que se valen de una neurología que postula ya corrientes eléctricas y distintas regiones del cerebro con funciones independientes. Por ejemplo, James sostiene que el objetivo de su artículo de 1884 es mostrar si hay un solo centro receptor de las emociones en el cerebro (como en el caso de Descartes, que sería la glándula pineal), o si las emociones son la conjunción de distintas reacciones en el cerebro (1884, p. 188). En el caso de Prinz, estos avances son mucho mayores. Durante el siglo xx, los avances en neurociencias y sus aplicaciones a la psicología son innu-

[28] Traducción libre del texto original: "In the tradition of Descartes, James, Lange, and, more recently, Damasio, I believe that emotions are inner estates that occur in response to bodily changes".

merables, de modo que no solo avanzó nuestra concepción del cerebro, sino también la misma forma de investigar en psicología y la manera en que entendemos nuestra vida mental. En cualquier caso, vemos que el corazón del programa versa sobre una descripción fisiológica de nuestras emociones. En otras palabras, entender nuestras emociones es necesariamente entender nuestro cuerpo; esto es lo que Prinz llama la “hipótesis de la corporeidad”.

El cinturón protector y la teoría de la evolución. La metodología de Lakatos, además de formular un núcleo firme, nos invita a ver en las teorías una serie de movimientos que permitan el progreso del programa en términos de predicciones y explicaciones novedosas. Vimos ya, muy brevemente, que la manera de entender el cuerpo y en especial el cerebro hace parte de estos movimientos. Pasar de una concepción del cuerpo como una máquina hidráulica, a la manera de Descartes, a una máquina eléctrica, como sucede en James y en Prinz, acarrea grandes progresos del programa. Por ejemplo, se hace posible dar cuenta con más detalle del papel de distintas regiones del cerebro en nuestras emociones, entre otros.

Entre estos movimientos quiero resaltar particularmente la introducción de la teoría de la evolución de Darwin que se da a partir de la teoría de James. Casos como este permiten ver cómo un programa de investigación puede nutrirse de otro; aquí, el programa sobre las emociones se nutre del programa sobre la biología. Como he expuesto, la teoría de Descartes atribuye a las emociones la función de llevar al alma a querer lo que es bueno y a evitar lo que

es malo. En otros términos, las emociones nos ayudan a preservar nuestro bienestar y alejarnos de lo que es nocivo. Descartes hace recaer estas funciones en la naturaleza, pero queda la pregunta sobre cómo han surgido. El autor le atribuye toda función natural a la creación de Dios. Él ha creado el mundo con leyes y le ha otorgado funciones al cuerpo para su preservación.

Desde James la explicación de la función de las emociones puede recaer, ya no en Dios, sino en la evolución de las especies. Al contar con la teoría de Darwin, James puede separarse de la idea de un plan divino y remitir la explicación de la función de las emociones a ventajas para la supervivencia y los mecanismos de selección natural. Eso hace eco de la declaración de Francisco Ayala:

> La mayor contribución de Darwin a la ciencia es que él completó la revolución copernicana al dibujar para la biología la noción de la naturaleza como un sistema de materia en movimiento gobernado por leyes naturales. Con el descubrimiento de la selección natural de Darwin, el origen y las adaptaciones de los organismos fueron traídas al ámbito de la ciencia. Los rasgos adaptativos de los organismos podían ahora ser explicados, como los fenómenos del mundo inanimado, como el resultado de procesos naturales, sin el recurso de un Diseñador Inteligente (Ayala, 2007, p. 3).[29]

---

[29] Traducción libre, el texto original dice: "Darwin's greatest contribution to science is that he completed the Copernican Revolution by drawing out for

Sin la necesidad del recurso de Dios, James nos permite hacer recaer en la labor científica las explicaciones acerca de la función de las emociones. Retomando los términos de Lakatos, hay aquí una forma de progreso teórico, en tanto que se predicen y explican nuevos hechos.

Además de la idea de que las emociones sirvan a la supervivencia, la teoría de Darwin sirve también para explicar las expresiones faciales de las emociones, como mostrar los dientes al sentir ira (como rezago de un gesto para ahuyentar una amenaza en otros animales). Incluso, podemos decir que la teoría de Darwin nos permite explicar parte de nuestras emociones a partir del estudio de animales no humanos, como lo que acabamos de mencionar.[30] Otro caso interesante de lo anterior, por ejemplo, es el de la evidencia empírica que Prinz asume a favor del programa naturalista. Al discutir la evidencia recopilada por Zajonc el autor acepta como argumento a favor de su teoría el hecho de que las emociones estén presentes en infantes y animales no humanos, esto es, que son ontogenética y filogenéticamente anteriores

biology the notion of nature as a system of matter in motion governed by natural laws. With Darwin's discovery of natural selection, the origin and adaptations of organisms were brought into the realm of science. The adaptive features of organisms could now be explained, like the phenomena of the inanimate world, as the result of natural processes, without recourse to an Intelligent Designer".

[30] Este movimiento resulta vital para el problema que trataremos en lo restante del texto. Parte de la naturalización de la intencionalidad que requieren las teorías naturalistas de la emoción pasa por notar que, si las emociones representan algo, esto es, si son representaciones mentales, estas deben tener una función. Esta función debe ser atribuida, al menos en principio, por un proceso evolutivo. De ahí que la inclusión de la teoría de Darwin no sea menor en este programa de investigación.

a la cognición —una explicación de tipo evolucionista— (Prinz, 2004b, p. 33-49). Esto significa que si las emociones tienen funciones anteriores a las de la cognición, entonces las funciones que estas cumplen recaen en la ventaja que las mismas supongan para la supervivencia y no requieren de juicios ni estructuras cognitivas que medien. En este caso, la descripción en términos fisiológicos puede valerse de la noción de primacía ontogenética y filogenética de la teoría de la evolución como una explicación válida sobre las emociones. Así, la introducción de la teoría de Darwin permite la expansión del programa de investigación.

Un problema más para el núcleo firme. Además de esto, vemos que la explicación que se quiere dar puede prescindir, de una u otra manera, de la cognición. En Descartes, las emociones son movimientos de la glándula pineal causados por el flujo de espíritus animales. Prinz, como Hatfield (2007a), sostiene que Descartes requiere del pensamiento para dar cuenta de las emociones, y por tanto no puede separarse completamente de la cognición (Prinz, 2004b, p. 11). No obstante, como señala Descartes en el *Tratado del hombre* —y como reconoce Hatfield (2007b)—, la descripción en términos mecanicistas puede darse tal y como se describiría a un autómata que es posible concebir sin alma (AT XI, 202). En este orden de ideas, las emociones funcionan como otro mecanismo de la máquina y no requieren necesariamente la inclusión del alma o de la cognición. En el caso de James, no solo es posible separar las emociones de la cognición, sino que es necesario, toda vez que involucrar la cognición sería una confusión. James argumenta constantemente que

es posible dar cuenta de las emociones sin que haya ningún juicio, como en los casos de enfermedades mentales (1884, p. 197).

En Prinz hay incluso varios ataques a la idea de que las emociones requieran de la cognición. Algunos de estos se valen de evidencia empírica, como el caso de la discusión sobre el debate Zajonc/Lazarus mencionado anteriormente. En esta discusión Prinz se decanta por la evidencia de Zajonc en contra de las teorías cognitivistas. Entre la evidencia que cita encontramos el papel que cumplen ciertas estructuras subcorticales como la amígdala en las emociones, la posibilidad de inducir emociones a partir de drogas e incluso expresiones faciales, la primacía ontogenética y filogenética citada atrás, entre otras (Prinz, 2004b, p. 33-49). Al discutir esta evidencia Prinz sostiene que las teorías cognitivistas se ven debilitadas, de modo que pareciera ser que las emociones no requieren de la cognición.

En síntesis, el programa naturalista busca una separación de la cognición para describir las emociones. Este rechazo al cognitivismo supone, sin embargo, un problema que el programa naturalista debe enfrentar. Las emociones, en principio, son intencionales; mis emociones son sobre algo. No obstante, mis reacciones fisiológicas no son sobre algo, como hemos señalado anteriormente. Retomando el ejemplo que he propuesto unas páginas atrás, la aceleración de mi corazón y la dilatación de mis pupilas no son sobre algo. En este orden de ideas, la descripción fisiológica no da cuenta de la intencionalidad de las emociones.

Este problema, el de dar cuenta de la intencionalidad de las emociones desde una perspectiva fisiológica (naturalizada), surge al querer describir, en términos físicos (no-intencionales) lo que, de suyo, es intencional. Como puede anticiparse, esta es una forma del problema tradicional de la naturalización de la intencionalidad. Dado que este se deriva del núcleo del programa naturalista, podemos intentar ver cómo las tres teorías en cuestión lo intentan solucionar. Entre ellas solo la teoría de Jesse Prinz involucra una teoría explícita de la intencionalidad. Sin embargo, las teorías de James y Descartes también intentan acomodar la intencionalidad a su explicación. Este será el objeto del siguiente capítulo, en el que mostraré con más detalle en qué consiste el problema y por qué resulta recalcitrante para una explicación naturalizada de las emociones.

2. La naturalización de la intencionalidad y el caso de las emociones

En el capítulo anterior hemos caracterizado lo que llamaremos el programa de investigación naturalista, el cual incluye las teorías de las emociones de Descartes, James y Prinz. En esta caracterización identificamos como núcleo central la definición de las emociones como la percepción de cambios en el cuerpo (siguiendo la formulación de James) y hemos adelantado algunos asuntos relativos al progreso teórico del programa, entre ellos, la introducción de la teoría de la evolución de Darwin y cómo esta permite acopiar evidencia biológica. Finalmente, ha quedado anunciado que la definición de las emociones que está en el corazón del programa naturalista supone un problema filosófico, a saber, el de la naturalización de la intencionalidad. En este capítulo quiero defender con cuidado esta última idea.

Puntualmente, quiero sostener que identificar las emociones con la percepción de cambios corporales requiere encontrar la relación entre estos cambios y los objetos de la emoción, esto es, entre un sustrato físico (corporal) y los elementos intencionales que hacen parte de nuestras emociones. En otras palabras, lo que se necesita es una descripción naturalizada de la intencionalidad de las emociones. Para

ello expondré primero de qué trata la noción de intencionalidad y por qué esta resulta filosóficamente problemática, en especial para la construcción de una ciencia natural sobre lo mental. Luego expondré algunas aproximaciones que pueden encontrarse en Descartes, James y Prinz sobre dicho problema. De este modo, podremos ver más adelante por qué el programa naturalista entraña algunas confusiones en su núcleo a partir de algunas de las observaciones de Wittgenstein.

2.1. La intencionalidad de las emociones

La tesis de Brentano. En la literatura contemporánea se le atribuye a Franz Brentano (1874/2009) la introducción (o reintroducción)[31] del concepto de intencionalidad en las discusiones sobre la mente. La importancia de Brentano reside en su idea de identificar la intencionalidad como el rasgo distintivo de los estados mentales (por contraste con los estados físicos). Con ello, Brentano busca delimitar el campo de estudio de la psicología como ciencia de lo mental. Como veremos más adelante, este protagonismo de la intencionalidad nos lleva a algunos problemas a la hora de explicar la vida mental en términos físicos.

La definición (o tesis, como ha sido llamada generalmente) de Brentano versa de la siguiente manera:

[31] Digo 'reintroducción', toda vez que la idea de que nuestros estados mentales tienen el rasgo de la intencionalidad ya estaba presente entre los escolásticos. Como veremos a continuación, Brentano reconoce esto, por lo que su tesis sobre la intencionalidad es un rescate más que un descubrimiento nuevo.

> Todo fenómeno mental está caracterizado por lo que los escolásticos de la Edad Media han llamado la inexistencia intencional (o mental) de un objeto, y que podríamos llamar, aunque no totalmente sin ambigüedad, la referencia a un contenido, la dirección hacia un objeto (que no debe ser tomado aquí como significando una cosa), o la objetividad inmanente. Todo fenómeno mental contiene en sí mismo algo como su objeto, pero no todos lo hacen de la misma manera. En la representación hay algo representado; en el juicio hay algo afirmado o negado; en el amor, amado; en el odio, odiado; en el deseo, deseado, etc. (Brentano, 1874/2009, p. 68).[32]

Según Brentano, entonces, todo fenómeno mental se caracteriza por hacer referencia a un contenido externo a sí mismo. Es importante señalar que este contenido, este objeto, podría no existir. Puedo, por ejemplo, pensar en un unicornio, caso en el cual mi pensamiento tiene un objeto (el unicornio) pero este no existe. De ahí que Brentano hable de la "inexistencia intencional". No obstante esta inexistencia,

---

[32] Traducción libre del autor, el texto original dice: "Every mental phenomenon is characterized by what the Scholastics of the Middle Ages called the intentional (or mental) inexistence of an object, and what we might call, though not wholly unambiguously, reference to a content, direction toward an object (which is not to be understood here as meaning a thing), or immanent objectivity. Every mental phenomenon includes something as object within itself, although they do not all do so in the same way. In presentation something is presented, in judgement something is affirmed or denied, in love loved, in hate hated, in desire desired and so on".

el estado mental (el pensamiento) sigue teniendo referencia a un contenido, esto es, sigue teniendo intencionalidad.

Cabe anotar, para lo que veremos más adelante, que el rasgo de la intencionalidad no es exclusivo de los estados mentales, como pensaba Brentano. Las palabras y oraciones en nuestro lenguaje también parecen tener carga intencional, toda vez que también refieren a algo por fuera de sí mismas. Por ejemplo, la palabra 'caballo' refiere a animales cuadrúpedos con tal o cual conjunto de características; la oración "Un caballo corre" refiere a un estado de cosas en el cual el animal se desplaza de tal o cual manera a tal o cual velocidad. En ambos casos, la palabra o la oración en cuestión tienen por objeto algo externo a ellas mismas, a saber, un animal o el desplazamiento de este. De ahí que la noción de intencionalidad sea muy cercana con la noción de significado. Lo que hemos identificado aquí como la referencia de una palabra a un objeto (caballo) o de una oración a un estado de cosas (el caballo corriendo) es, en principio, lo que esta palabra u oración significan. Por ello, la preocupación por la intencionalidad corre de la mano con la preocupación por la semántica. En otras palabras, explicar cómo una palabra puede significar es, cuando menos, análogo a explicar cómo un estado mental puede referir a un objeto.

La intencionalidad de las emociones. En un sentido muy general, es fácil ver cómo las emociones pueden tener carga intencional. Cuando digo "Estoy feliz por el ascenso que recibí en el trabajo", puedo decir que el objeto de mi felicidad es el ascenso. Además, es posible que este objeto (en este caso, mi ascenso) no se haya dado, es posible que

tenga el rasgo de la "inexistencia intencional" que sugería Brentano, y que realmente no me hayan ascendido aunque yo crea que así fue. Ahora, podemos preguntarnos, ¿será posible aplicar esta tesis a todo el programa naturalista sobre las emociones?

En el caso de Descartes, el autor sí parece tener una noción de lo que podríamos llamar "fuerza intencional" de nuestros estados mentales. En las *Reglas para la dirección del espíritu* (1628/1996), en particular en la Regla XII, el autor dice que "la fuerza por la cual propiamente conocemos las cosas es puramente espiritual" (AT X 415). Esta fuerza determina cómo nos dirigimos a los objetos de los que nos ocupamos. Dice Descartes:

> Y es una sola y misma fuerza, la cual, si se aplica con la imaginación al sentido común, es denominada ver, tocar, etc.; si se aplica a la imaginación sola en cuanto ésta está revestida de diversas figuras, es denominada recordar; si a la imaginación para formar nuevas figuras, decimos imaginar o concebir; si finalmente, actúa sola, entender (AT X 415).

Tenemos aquí que esta fuerza determina cómo nos ocupamos de los objetos que se presentan ante la mente y que, dependiendo de cómo esta fuerza haga uso de la imaginación y el sentido común, dará lugar a distintos estados mentales. Por ejemplo, cuando esta fuerza se ocupa de la imaginación por sí sola la llamamos "entender"; pero cuando se ocupa de la imaginación y de las figuras que esta recibe

del sentido común[33] la llamamos "ver" o "tocar", etc. En cualquier caso, hay una fuerza que determina lo que podemos llamar la "carga intencional" de un estado mental. Esta fuerza se aplica a los objetos de mis estados mentales —de ahí que mis estados sean intencionales, es decir, que tengan contenido— y cuando esta fuerza se aplica de una u otra manera, decimos que cambia el estado mental en el que me encuentro (imaginar, entender, ver, tocar, etc.). Esto, en principio, parece permitir una aproximación a la tesis de Brentano desde la teoría cartesiana.

Además de esto, el proyecto científico cartesiano viene acompañado de una suerte de noción de representación mental. En el primer capítulo del *Tratado del hombre*, dice Descartes:

> Al ofrecer una explicación de la luz, lo primero sobre lo que deseo llamar su atención es en que es posible que haya una diferencia entre la sensación como la tenemos, esto es, la idea que formamos de ella en nuestra imaginación a través del intermediario de nuestros ojos, y lo que está en los objetos que produce esa sensación en nosotros, esto es, lo que está en la llama o el Sol que llamamos 'luz'. Pues aunque todos están comúnmente convencidos de que las ideas que tenemos en nuestro pensamiento son exactamente como los objetos de las que proceden, no

[33] En este caso, la expresión 'sentido común' significa algo cercano a 'sensación'.

> conozco ningún argumento convincente para esto (Descartes, AT X 3).

Si las cosas no nos parecen lo que son realmente, esto es, si hay una distinción entre los objetos en tanto objetos de nuestra mente y en tanto objetos reales, entonces pareciera que podemos hablar de una distinción entre representación y objeto. Nuestros estados mentales representan objetos gracias a ciertas conexiones causales con ellos, pero estos objetos no nos tienen que aparecer como son realmente. Esto permitiría una lectura de Descartes que pueda acercarse a la manera en que Dretske —y, consigo, Prinz— naturalizan la intencionalidad. Sobre esto volveremos más adelante.

Por el lado de la psicología de James, el tema de la intencionalidad aquí resulta bastante complejo. James nunca se refiere explícitamente al término 'intencionalidad', aunque en algunos casos sí usa el término 'intención'. Si bien sabemos también que James leyó a Brentano (PP, p. 187), el autor no es muy claro sobre su posición frente a la intencionalidad de nuestros estados mentales. Además, el autor niega que las emociones involucren elementos cognitivos, lo que sugiere que podrían no tener el carácter de la intencionalidad. No obstante, abordaré la teoría de James desde una lectura que muestre que para el filósofo las emociones sí son intencionales, al punto que el autor incluso intenta aclarar la noción de "objeto (intencional)" sobre el que versan las mismas.[34]

---

[34] Para otra lectura sobre la intencionalidad de las emociones en la teoría de James, véase Ratcliffe (2005) y Wilshire (1969). Me plegaré a lecturas simi-

En principio, James acepta la tesis de Brentano y habla de una distinción entre un estado mental y su objeto. A lo largo de su psicología y de su teoría de las emociones la distinción aparece una y otra vez. Incluso, al final de su artículo de 1884, al describir la fisiología cerebral de las emociones, James sostiene que el cerebro percibe un objeto que luego transforma en objeto-aprehendido y objeto-emocionalmente-sentido (James, 1884, p. 203). Así, estos objetos parecen tener carga intencional; en unos casos son aprehendidos y en otros están acompañados de una emoción.

No obstante, hay aclaraciones pertinentes sobre la noción de objeto. En el capítulo sobre "El flujo del pensamiento", James sostiene que el objeto del pensamiento no es un objeto completamente aislable. El ejemplo del autor es el del pensamiento "Colón descubrió América en 1492". En principio, el objeto de mi pensamiento es Cristóbal Colón y su descubrimiento. Sin embargo, James advierte que esto no es preciso. El objeto de mi pensamiento es, para James, el objeto en sus relaciones totales. Así, el autor sugiere que escribamos el objeto de este pensamiento como "Colón-descubrió-América-en-1492". Esto significa que no pensamos el objeto del pensamiento solo como Colón, o su descubrimiento, sino todas las relaciones entre Colón y la manera en que se está pensando (PP, p. 275). Para dar otro ejemplo, pensemos en la percepción de una mesa. Para

lares, en donde la teoría de James acepta la intencionalidad de las emociones; sin embargo, no discutiré este aspecto de la interpretación de James, toda vez que mi interés yace en la tesis central del programa de naturalización de las emociones y no en la exégesis de un autor particular.

James, el objeto de mi estado mental no es solamente la mesa; es la mesa en su contexto, con todo y relaciones de espacio, ambiente, iluminación, etc. En palabras del autor:

> El objeto de cada pensamiento, entonces, es nada más y nada menos que todo lo que el pensamiento piensa, exactamente como el pensamiento lo piensa, cuan complicado pueda ser el asunto, y cuan simbólica la manera de pensarlo pueda ser (PP, p. 276).[35]

Tenemos entonces que, para James, los objetos de nuestros estados mentales son objetos tomados en su totalidad, en todas sus relaciones tal y como se dan en la conciencia. Esto también aplica para las emociones. En las aclaraciones y respuestas de James en 1894 sobre su teoría de las emociones, el autor afirma:

> Los 'objetos' son ciertamente los disparadores primitivos de movimientos de reflejo instintivos. Pero ellos tienen lugar, mientras la experiencia transcurre, como elementos en 'situaciones' totales […]. Tan pronto como un objeto se ha vuelto entonces familiar y sugestivo, sus consecuencias emocionales, *en cualquier teoría de la emoción*, deben comenzar mejor en la situación total que sugiere más que de su presencia desnuda (James, 1894/1994, p. 206; énfasis en el original).

[35] Como señala Bruce Wilshire, esto puede tener cierto parecido con la manera en que Husserl trata los objetos del pensamiento. Véase Wilshire (1969).

Esto invita a introducir varias aclaraciones. Primero, tenemos que los objetos de la emoción se dan en situaciones totales, como advierte también el autor sobre la conciencia en los PP. Estos objetos también son objetos en tanto aprehendidos por la percepción (objeto-aprehendido) y en tanto objetos de una emoción (objeto-emocionalmente-sentido). Sin embargo, en principio, la distinción de Brentano se mantiene, aunque con algunos ajustes.

Finalmente, atendamos a la teoría de Prinz y cómo ella maneja la noción de intencionalidad. Para ello debemos dejar entrar primero una distinción que no hemos mencionado hasta el momento, esta es, la distinción entre "objeto formal" y "objeto material". Para Prinz, las emociones tienen por *objeto formal* un tema relacional básico, esto es, una relación con el ambiente, mientras que tienen por *objeto material* el objeto que provoca mi emoción. ¿Qué significan las expresiones «objeto formal» y «objeto material» aquí? Anthony Kenny (1963/2003) desarrolló —o retomó— esta distinción con la que ya contaban los escolásticos al hablar de intencionalidad. La distinción mencionada entre objetos formales y materiales puede resumirse de la siguiente manera: "Cualquier cosa que pueda ser ϕ-ado es el objeto material de ϕ-er"[36] (Kenny, 1963/2003, p. 132). Si el unicornio del que hablábamos antes puede ser imaginado, entonces el unicornio es el objeto material de mi imaginación de unicornio. Del mismo modo, en el caso de la felicidad por mi

[36] Traducción libre, el texto original dice: "Anything which can be ϕd is a material object of ϕing".

ascenso, el ascenso es el objeto material de mi felicidad. Por otra parte, un objeto formal es "el objeto bajo esa descripción [la del estado mental en cuestión] que *debe* aplicarse a él si es posible ϕ-arlo"[37] (Kenny, 1963/2003). Si digo que el unicornio es imaginado, entonces "el objeto imaginado" es el objeto formal de mi imaginación; en otras palabras, mi imaginación es sobre lo imaginado, y no podría ser sobre otra cosa. Esto, por supuesto, puede sonar trivial. Sin embargo, hay descripciones de objetos formales que no son triviales. Tomando el ejemplo de Kenny, solo algo sucio puede ser objeto de la limpieza, i.e., la limpieza es sobre lo sucio. Esto nos informa algo sobre lo que significa limpiar, sin especificar un objeto materia. Puedo limpiar mi escritorio, en cuyo caso el escritorio es el objeto material de la limpieza y por tanto podemos inferir que el escritorio ha de estar sucio para poder ser limpiado.

En el caso de las emociones, Prinz sostiene que los objetos formales son temas relacionales básicos (2004b, p. 68). Pensemos nuevamente en el ejemplo de la felicidad por mi ascenso. Ya hemos dicho que el objeto material de mi felicidad es mi ascenso. En la tabla de temas relacionales básicos a la que acude Prinz (2004b, p. 16), la felicidad tiene como tema "hacer un progreso razonable hacia la realización de una meta".[38] Esto significa que mi felicidad siempre es sobre

---

[37] Traducción libre, el texto original dice: "is the object under that description which must apply to it if it is to be possible to ϕ it".

[38] Traducción libre, el texto original dice: "Making reasonable progress toward the realization of a goal".

"aquello que representa progreso hacia mis metas" —así como mi imaginación siempre es sobre lo imaginado—; dado que mi ascenso representa tal progreso, este puede ser el objeto material de mi emoción. Nótese que la conexión entre la imaginación y lo imaginado es una conexión lógica, no contingente (mi imaginación no puede ser sobre algo además de lo imaginado). Del mismo modo ocurre con estos temas relacionales básicos. Mi felicidad siempre es sobre aquello que representa progreso, de modo que cosas como encontrarme una peligrosa serpiente no serían objeto de mi felicidad, salvo si —nuevamente— encontrar una serpiente me acercase a una meta. La serpiente sería, en cambio, objeto de mi temor, toda vez que el temor tiene por objeto formal "aquello que representa peligro".[39]

De lo avanzado hasta ahora podemos ver que en el programa naturalista las emociones podrían verse como intencionales, esto es, tienen un objeto al cual se refieren. Además de la versión de Brentano, que podemos intentar ver en Descartes y en James, Prinz introduce la distinción entre objetos formales y materiales para refinar aún más la noción de intencionalidad para las emociones. Sin embargo, el hecho de que las emociones sean intencionales introduce un problema clásico en filosofía. Si las emociones, entendidas como reacciones corporales, son intencionales, ¿qué da cuenta de su contenido o de su objeto? En principio, las reacciones corporales no son intencionales por sí mismas.

[39] Para ver la lista completa véase Prinz (2004b, p. 16).

¿Cómo podemos explicar entonces este rasgo? Aclaremos este problema.

2.2. El problema de la intencionalidad

El proyecto naturalista de lo mental. En las últimas décadas se ha discutido ampliamente sobre la posibilidad de describir estados mentales con lenguaje no intencional, especialmente en lenguaje fisicalista. La idea es, *mutatis mutandis*, lograr unidad en las ciencias, de modo que pueda adelantarse una reducción de la psicología a una ciencia básica con un lenguaje unificado. Por lo general, es la física la que asume el papel de ciencia básica, y la reducción se convierte entonces en una reducción fisicalista. A este proyecto, y sus posibles variaciones (como las reducciones no fisicalistas), se ha denominado "proyecto naturalista" o "proyecto de naturalización". Un esquema aproximado del naturalismo en ciencia es ofrecido por Steven Horst (2007):

> *Naturalismo—un esquema general:* el naturalismo sobre el dominio D es la posición según la cual todas las propiedades de D han de ser acomodadas dentro del marco de la naturaleza como es entendida por las ciencias naturales (Horst, 2007, p. 13).[40]

[40] Traducción libre del autor, el texto original dice: "Naturalism –a General Schema: Naturalism about domain D is the view that all features of D are to be accommodated within the framework of nature as it is understood by the natural sciences".

Según este esquema, el naturalismo (o el proyecto naturalista) en el dominio de la psicología implica acomodar las propiedades psicológicas (o predicados psicológicos) al marco de la naturaleza y este marco, como dije anteriormente, es supuesto en general como un marco fisicalista. Esta acomodación al marco de la física tiene como ventajas, entre otras, la unificación de una misma ontología —una ontología en la que solo hay objetos físicos en el mundo—, y la unidad en la explicación (la construcción de explicaciones y leyes se apoya en las de la física).

Este proyecto, por supuesto, ha sido objeto de extensos debates[41] durante los últimos años y es un punto de discusión que continúa vigente. Por lo pronto me interesa exponer el proyecto de naturalización desde una óptica fisicalista. Así, espero mostrar cómo este proyecto hace parte nuclear del programa de investigación sobre las emociones que estamos estudiando.

El problema de la representación mental. Tenemos, entonces, que el proyecto de naturalización de lo mental busca acomodar, o reducir, lo mental a lo físico. De esto se sigue, volviendo a la definición de Brentano, que acomodar

---

[41] No todos aceptan la posibilidad de la naturalización de lo mental. Por ejemplo, Davidson (1970/2002) no acepta que puedan construirse leyes y explicaciones físicas de eventos mentales (leyes psicofísicas). Sobre el aspecto ontológico, Woodruff Smith (1999/2004) intenta construir una ontología en la que no solo haya objetos físicos, sino también intencionales. En cualquier caso, ambas ventajas no son del todo aceptadas y soy consciente de la discusión alrededor de ellas. No obstante, mis propósitos en este capítulo son, como he sugerido, expositivos y no quiero comprometerme con una posición particular. Más adelante entraré a discutir este proyecto, pero por lo pronto deseo simplemente presentarlo.

los eventos mentales en el mundo físico implica explicar, en términos fisicalistas, qué es la intencionalidad —que, por lo pronto, es el rasgo distintivo de los estados mentales—. En otras palabras, si la reducción de lo mental a lo físico es posible, debe ser posible explicar cómo de algo físico puede surgir el extraño fenómeno de que mis estados mentales tengan un contenido, que refieran a algo por fuera de ellos. Sin embargo, hay varios argumentos para ver cómo una descripción física parece quedarse corta para dar tal explicación. Tomemos, por ejemplo, el caso de la representación, citando un famoso argumento de Hilary Putnam (1981/2006). El autor propone el siguiente experimento mental:

> Una hormiga se arrastra lentamente sobre la arena. Conforme avanza va trazando en ésta una línea. Por puro azar, la línea se desvía y vuelve sobre sí misma, de tal forma que acaba pareciendo una reconocible caricatura de Winston Churchill. ¿Ha trazado la hormiga un retrato de Winston Churchill, un dibujo que *representa* a Winston Churchill? (Putnam, 1981/2006, p. 15; resaltado del texto).

Como advierte Putnam, intuitivamente no diríamos que la hormiga ha dibujado una representación de Winston Churchill. La hormiga no conoce a Winston Churchill, mucho menos su imagen, ni cómo dibujarla. Frente a este caso, decimos que la hormiga no ha tenido la *intención* de dibujar a Winston Churchill y, por lo tanto, que la imagen no califica como una representación. En cambio, si un artista hiciera el mismo dibujo en la arena, este sí fungiría como

una representación, toda vez que el artista lo ha cargado con intencionalidad (i.e., le ha dado un significado, de modo que el dibujo refiere al personaje británico). El artista, a diferencia de la hormiga, conoce a Winston Churchill y su traza resulta representándolo.

Ahora bien, la descripción física del dibujo de Winston Churchill puede ser la misma para ambos casos (el del artista y el de la hormiga). Es perfectamente posible que los granos de arena queden dispuestos de la misma manera tras ambas trazas, de modo que todas sus coordenadas espaciotemporales sean exactamente iguales. Sin embargo, al menos de manera intuitiva, uno de los casos es una representación (intencional) y el otro no lo es. De allí se sigue que la descripción física no baste, en principio, para que haya intencionalidad (o significado).

Lo mismo que funciona aquí para una representación pictórica puede extenderse (como también hace Putnam) a los estados mentales. Si nuestros estados mentales son equivalentes a estados cerebrales (físicos), ¿qué da cuenta de que un estado mental represente algo, toda vez que la descripción física no basta? Siguiendo de la mano con Putnam, pensemos ahora en un grupo de humanos extraterrestres (con la misma disposición cerebral que nosotros) que jamás han visto ni oído hablar de un árbol. Supongamos ahora que en su planeta cae de repente un dibujo de un árbol. Cuando ellos ven el dibujo, tienen un estado mental, en este caso la percepción del dibujo. No obstante, su estado mental no es el de percibir una representación de un árbol. Es, a lo sumo, un estado que representa una serie de manchas que es

probable que carezcan de sentido. En otras palabras, ellos no ven propiamente un árbol; ven manchas. Sin embargo, su estado cerebral es, por hipótesis, exactamente idéntico al que tendríamos nosotros si viéramos este dibujo, aunque nosotros sí diríamos que estamos viendo un árbol (o el dibujo de un árbol). Luego, la disposición cerebral, por sí sola, no da cuenta de la intencionalidad de nuestro estado mental, i.e., del hecho de que nuestro estado mental tenga por objeto una mancha o un árbol.

Este argumento intenta ilustrar, entonces, cómo la descripción física de un fenómeno (palabra, oración, estado mental, etc.) no basta para dar cuenta de su intencionalidad. En otras palabras, la referencia a un objeto o el contenido no son explicados por la descripción física. Esto, por supuesto, pone en jaque el proyecto de naturalización, toda vez que parecería que hay algo que escapa a la física. Lo mismo ocurre con los intentos por naturalizar las emociones que hemos tratado. Al intentar explicar las emociones únicamente a partir de reacciones corporales nos topamos con la dificultad de que una explicación fisiológica no basta para explicar la intencionalidad. Podemos elaborar un caso análogo al expuesto por Putnam para aclarar esta cuestión. Imaginemos nuevamente a los extraterrestres con nuestra misma disposición cerebral y que ahora ven caer un dibujo, ya no de un árbol, sino de la trágica muerte de Julieta del drama de Shakespeare. Es concebible que estos extraterrestres, al no haberse enterado jamás de la obra de Shakespeare, tengan exactamente las mismas disposiciones corporales, esto es, que su corazón se acelere y sus ojos se

tornen llorosos, solo que en ese caso diríamos que sienten tristeza por la muerte de una persona y no por alguna razón relacionada con la historia de Romeo y Julieta. El contenido de su emoción es diferente, incluso teniendo, en principio, los mismos estados corporales. De ahí que la descripción en términos de estados corporales pareciera ser insuficiente para explicar por qué ellos sienten tristeza por algo diferente a por lo que lo hacemos nosotros.

Algunos piensan que este problema de dar una descripción naturalizada de fenómenos mentales como la intencionalidad resulta crucial para nuestra comprensión de los conceptos psicológicos mismos. En particular, parece que la imposibilidad de naturalizar un término puede resultar en una posición eliminativista, es decir, puede usarse para defender la idea de que debemos exiliar los términos mentales del vocabulario de la ciencia. Esta posición se atribuye especialmente a Quine (1960) y ha sido defendida con mayor fuerza por Churchland (1981). Moya (1994) resume el problema de la siguiente manera:

> Lo que Quine nos ha hecho ver con claridad es que la autonomía e irreductibilidad de las propiedades intencionales es un arma de doble filo. Puede ser usada tanto para defender la importancia y sustantividad de las propiedades mentales que en la vida ordinaria atribuimos a los seres humanos como para apoyar la futilidad e inexistencia última de dichas propiedades y del supuesto conocimiento interpretativo que basamos en ellas (Moya, 1994, p. 228).

Esta amenaza ha llevado a varios intentos por naturalizar la intencionalidad que buscan salirle al paso a los obstáculos mencionados. Más adelante intentaré mostrar que tal amenaza se funda en una mala comprensión de la gramática de ciertas preguntas, de modo tal que no tenemos que ver el problema de la intencionalidad como un misterio que debamos resolver como problema filosófico.

No obstante, primero debemos tener más claridad sobre cómo el programa de naturalización de las emociones ha llevado a cabo la tarea de explicar la intencionalidad de estas. Me interesa, muy especialmente, el programa de Fred Dretske (1995) y su teoría de la representación mental. En el capítulo anterior mencioné que Prinz, para solucionar el problema de la intencionalidad que no lograba acomodar James, hace uso de esta teoría de la representación. Expondré entonces, brevemente, aunque con algo más de detalle que en el capítulo anterior, en qué consiste dicha teoría, para luego mostrar su acomodación e importancia para una teoría naturalista de las emociones. Sin embargo, veamos primero cómo Descartes y James, siendo anteriores a Dretske, podrían intentar superar la dificultad que supone la intencionalidad de las emociones en una explicación naturalizada.

2.3. Aproximaciones al problema de la intencionalidad en el programa naturalista

Representación y naturalización en la ciencia cartesiana. Hemos dicho que en el proyecto científico cartesiano se ofrece una explicación mecanicista del mundo, esto es, una

descripción en términos causales —en contraposición a las explicaciones teleológicas aristotélicas—. En esta descripción, el cuerpo humano aparece como una máquina que es afectada por —o hace parte de— la gran máquina que es el mundo. Además de esto, hemos visto cómo en Descartes parece haber una noción de representación. El autor abre el *Tratado de la luz* afirmando que la manera en la que los objetos se nos presentan no debe parecerse a la manera en que ellos son. Podemos distinguir entre las propiedades de los objetos y la manera en que los vemos. El análisis causal debe explicar qué sucede en ellos que permite que los sintamos, sin que esto signifique que debamos percibir todas sus propiedades.

Aclaremos mejor esta noción de representación. En la misma apertura del *Tratado de la luz* Descartes arroja un argumento con grandes parecidos con el desarrollo del argumento de Putnam citado anteriormente. Recordemos que el argumento de Putnam se basaba en la idea de que la cara de Winston Churchill dibujada por una hormiga no era en sí misma una representación de Winston Churchill. Putnam desarrolla este argumento para aplicarlo también al lenguaje: si en lugar de la cara, la hormiga hubiese trazado la palabra “CHURCHILL”, esta palabra igualmente no referiría al personaje británico. Algo similar dice Descartes:

> Sabéis perfectamente que las palabras, sin tener ningún parecido con las cosas que significan, nos permiten concebirlas e incluso, a menudo, sin que nos apercibamos del sonido de los términos ni de sus sílabas [...] Pues bien,

> si las palabras, que solo significan por institución de los hombres, bastan para hacernos concebir cosas con las que no tienen ningún parecido ¿por qué la naturaleza no podía también haber establecido cierto signo que nos produzca el sentimiento de la luz, a pesar de que este sigo no tenga nada en sí que sea parecido a este sentimiento? (AT XI 4).

Analicemos este pasaje. Descartes —como Putnam— sostiene que las palabras (su sonido, sílabas, etc.) no se parecen a los objetos a las que ellas refieren. La traza "CHURCHILL" en la arena no se parece a Winston Churchill, incluso la imagen trazada tampoco tiene una semejanza total. Lo que hace que estas trazas adquieran significado o un objeto es una carga intencional que ponemos sobre ellas. Aquí conviene retomar la "fuerza del espíritu" de la que hablaba Descartes en las *Reglas* y que denominamos "fuerza intencional".

Lo mismo ocurre con nuestras sensaciones y otros estados mentales. Nuestros estados mentales, entre ellos las emociones, pueden *representar* objetos (signos) incluso si no se parecen a ellos. La naturaleza, dice el autor, ha instituido signos, i.e., propiedades en los objetos que los hacen susceptibles de provocar en mí una emoción u otro estado mental, incluso si la manera en que se me presentan los objetos no se parece a aquello que causa una reacción corporal en mí. Así las cosas, mi cuerpo tiene la función de detectar propiedades en los objetos y así provocar —mediante movimientos de la glándula pineal— reacciones en el alma. Para

dar un ejemplo, un objeto debe poder generar vibraciones en nuestro tímpano para que podamos escucharlo. Sin embargo, nosotros no sentimos propiamente vibraciones en el tímpano; sentimos un ruido (sea, digamos, la melodía de un violín). Podemos decir que nuestro cuerpo tiene la función de captar estas vibraciones y provocar en el alma el sonido del violín.

Ahora bien, ¿qué conexión existe entre mi sensación y el objeto que la produce? O mejor, ¿cómo puede mi sensación referir a un objeto? Ya podemos intuir una respuesta; la clave está en la conexión causal. El objeto tiene propiedades —o signos— que Dios ha instituido que provoquen en mí ciertas reacciones. En el caso de las emociones, cuando percibimos un objeto que provoca en nosotros una emoción (sea, por ejemplo, la criatura monstruosa de la que habla Descartes que provoca en nosotros temor) podemos decir que mi cuerpo reacciona causalmente a ese objeto, por un lado, con la simple percepción del objeto (movimientos en el ojo, el nervio óptico, etc.) y, por otro, con los cambios que son mi emoción (agitación del corazón, movimiento de los espíritus animales a los músculos, etc.). Así las cosas, decimos que tememos a la criatura monstruosa porque es ella la que ha excitado en mí estas reacciones. En otras palabras, en la teoría cartesiana decimos que mi temor es sobre la criatura monstruosa porque mi cuerpo tiene la función, dada por Dios, de detectar en la criatura ciertas propiedades que causan en mí ciertas reacciones corporales.

Pese a esta intuición, el análisis cartesiano solo permite llevar la cadena causal hasta la glándula pineal. No obstante,

la imagen (o la reacción) que se da en la glándula pineal no es la sensación que se presenta frente al alma. Hay allí un abismo y una gran ambigüedad sobre lo que ocurre entre la glándula pineal y el alma. Con todo, no podemos decir que hay una conexión causal de los objetos hasta el alma. Es aquí donde, parece, requerimos de una teoría de lo mental, una ontología diferente, que permita extender las relaciones causales o —cuando menos— que permita saldar el abismo entre mente y cuerpo.

James: naturalismo y holismo. En la filosofía de la psicología de James la noción de 'naturalización' tiene varias aristas que es importante tomar con cuidado. Para empezar, son varias las interpretaciones que podemos dar a la teoría de lo mental de James. Cooper (1990) identifica cuatro lecturas de ella en la literatura sobre el autor: el monismo neutral, el panpsiquismo, el naturalismo y la protofenomenología. Si bien no es objeto de este estudio decantarse por una de las alternativas (o una conjunción de varias, como la que presenta el mismo Cooper), es importante ver de antemano la gran variedad de interpretaciones que podemos adoptar. Para nuestros motivos, intentaremos mantener una neutralidad frente a la teoría de lo mental de James, esto es, intentaremos no adoptar una posición explícita, dado que esto sería motivo de otro trabajo. Así las cosas, examinaremos algunas recomendaciones de James y algunos de sus comentaristas para elucidar el estatus que tiene la intencionalidad en su teoría de las emociones.

Lo primero que hemos de aclarar es que la psicología, en la obra de James, aparece como la formulación de un

programa de investigación y no como una ciencia en funcionamiento. La idea de James es justamente formular de manera precisa cómo deberá proceder una ciencia natural que estudie lo mental. En este marco, James presenta una distinción entre dos niveles de investigación, uno científico y otro metafísico. En los PP, dice el autor:

> Cada ciencia natural asume ciertos datos acríticamente, y se niega a retar los elementos entre los que se obtienen sus mismas "leyes", y de los cuales se llevan a cabo sus propias deducciones. La psicología, la ciencia de las mentes individuales finitas, asume como sus datos (1) pensamientos y sensaciones, y (2) un mundo físico en el tiempo y el espacio en el que coexisten y que (3) ellas conocen. Por supuesto estos datos son de suyo discutibles; pero la discusión de ellos (como de otros elementos) es llamado metafísica y está por fuera de la provincia de este libro. Este libro, asumiendo que los pensamientos y las sensaciones existen y son vehículos de conocimiento, sostiene que la psicología, cuando ha afirmado la correlación empírica de varios tipos de pensamiento o sentimiento con condiciones definidas del cerebro, no puede ir más allá –no puede ir más allá, esto es, como una ciencia natural. Si va más allá, se vuelve metafísica (PP, p. vi).

De aquí podemos extraer varias tesis. Primero, la psicología —en tanto proyecto de ciencia natural— debe presuponer *(1)* la existencia de pensamientos y sensaciones,

(2) la existencia de un mundo físico con el que coexisten estos pensamientos y sensaciones y *(3)* que estos últimos (pensamientos y sensaciones) conocen, o mejor, interactúan, con el mundo. La psicología, entonces, no se ocupará de dar cuenta de dichos principios; estos son su punto de partida. Es la filosofía, o la metafísica, la que ha de ocuparse de ellos. Así las cosas, la filosofía es la que deberá aclarar la naturaleza metafísica de la vida mental y del mundo físico; la psicología solo deberá presuponerlos acríticamente.

Tenemos, entonces, que la psicología tiene estas tres tesis como presupuestos. Ahora bien, ¿qué debe hacer la psicología con ellos? Su labor será la de encontrar correlaciones entre nuestra vida mental y nuestros estados físicos. De manera resumida, la tarea de la psicología es la formulación de leyes psicofísicas. Para citar a James: "No es que al día de hoy tengamos una 'ciencia' de la correlación entre estados mentales con estados cerebrales; pero sí que la afirmación de las leyes de tal correlación forma el programa de una ciencia bien delimitada y definida" (1892, p. 151).[42] Si bien en esta formulación se habla de "estados cerebrales", James recuerda al lector constantemente en los PP que se trata, no solo del cerebro, sino de estados fisiológicos de toda índole (PP p. 5).

Ahora, ¿cómo se aplica esto a la naturalización de la intencionalidad? Hemos de anotar, primero, que James propone una psicología científica que no se ha de ocupar de querellas metafísicas como el estado ontológico de lo

---

[42] Traducción libre del autor.

mental (PP vi) o de la intencionalidad. Por lo tanto, la naturalización de la intencionalidad pareciera ser objeto de la metafísica, no de la psicología. La psicología ha de suponer que la intencionalidad deberá tener alguna cabida en el mundo natural; he ahí la postura naturalista de James.[43]

Sin embargo, cuando acudimos a la metafísica del autor, no encontramos claridad sobre este asunto. Podemos identificar, quizás, dos dificultades importantes. La primera es la forma en la que James entiende la intencionalidad de nuestros estados mentales. Ya hemos advertido anteriormente que para James la noción de 'objeto' que supone la estructura de la intencionalidad es un objeto en su totalidad, i.e., un objeto no es un objeto aislado para nuestra vida mental, sino un objeto en una situación con un sinfín de relaciones. En este sentido, dado este carácter holista —si se quiere— de lo mental, pareciera que no es posible dar un correlato unívoco en términos cerebrales de lo que constituye la intencionalidad. Notemos que una línea de argumentación similar es la de Davidson (1970/2002), en la que el carácter holista de lo mental sirve de defensa de la imposibilidad de construir leyes psicofísicas. Si bien Davidson no apela a estructuras intencionales como las entiende James —para Davidson el holismo no se encuentra en el objeto de nuestros estados mentales—, este tipo de argumentación puede mostrarnos

[43] Hemos dicho que Cooper identifica, entre otras, la posición naturalista como una lectura de la teoría de lo mental de James. Sin embargo, esta lectura es desarrollada, no por Cooper, sino por Flanagan (1991). Cooper argumenta que la lectura de Flanagan es útil al nivel científico, mas no al nivel metafísico, de la psicología de James.

que aceptar tesis holistas puede resultar problemático para una naturalización.

La segunda dificultad es la ontología misma que desarrolla el autor, en la cual —como sugiere Cooper— se cuantifica sobre experiencia pura y no sobre estados mentales o físicos (1990, p. 579). En otras palabras, lo que existe es la experiencia pura y de ella se construyen los objetos físicos (cuerpos) y los estados mentales; de ahí el título de "empirismo radical". Si bien no es objeto de nuestra investigación profundizar en esta tesis, es central anotar que esto dificulta la reducción ontológica que parece estar detrás del proyecto naturalista. Este proyecto supone no la experiencia pura, sino la ontología propia de la física. Para decirlo de otra manera, el proyecto naturalista, en tanto busca unidad de las ciencias, asume como base la ontología fisicalista y no una forma de empirismo radical como la de James.

No obstante estas dificultades, podemos ver que James sí intenta apostarle a una ciencia que ofrezca correlatos entre la vida mental y los eventos físicos, independientemente de la interpretación metafísica de ellos. Podemos decir, incluso, que James le apuesta a la unidad de las ciencias. Para citar al autor: "Para empezar, ¿qué es una ciencia natural? Es un mero fragmento de la verdad tomado de la masa entera de ella por mor exclusiva de la efectividad práctica" (1892, p. 147).[44] Así, toda ciencia se encarga de un trozo de la verdad, pero, en el fondo, todas deben remitir a esa misma verdad unificada.

---

[44] Traducción libre del autor.

Por ello podemos decir que James le apuesta a una psicología naturalizada. El científico, en tanto que no se ocupará de querellas metafísicas, presupondrá la existencia de un mundo exterior y de una mente, e intentará develar las correlaciones que existen entre ambos mundos. Su objetivo será encontrar leyes naturales que permitan explicar cómo el mundo físico causa eventos mentales y viceversa. En resumen, el psicólogo intentará acomodar a la mente en el dominio de lo natural. Notemos que la teoría de las emociones de James se inscribe en este proyecto. La manera en la que James formula su teoría es la de encontrar correlatos fisiológicos a nuestras emociones; de ahí que nuestras emociones sean, en último término, reacciones corporales.

Si bien esto es tan solo la formulación de un programa de investigación, nótese que define precisamente la agenda que luego dará lugar al proyecto de naturalización de la intencionalidad. Descartes ofrecía una explicación mecanicista de las pasiones, aunque de la mano con un dualismo; James, en cambio, no solo le apuesta a una explicación mecanicista, sino que formula una agenda que salde el abismo cartesiano. En otras palabras, James ve la distancia entre mente y cuerpo que supone la ciencia cartesiana y propone una psicología cuyo objetivo sea explicar esta brecha. Para decirlo en una frase, James nos ofrece una de las primeras formulaciones del proyecto de naturalización.

En el campo de las emociones, James sostiene, en los PP, que ya hay suficientes descripciones de nuestra vida emocional, y lo que nos interesa es su descripción causal (PP, pp. 448-449). En este aspecto decíamos que James busca

una explicación mecanicista. Ahora, ¿cómo explicar la intencionalidad de las emociones? Podríamos preguntarnos, ¿qué explica que ciertas emociones tengan ciertos objetos? Para James, como vimos en el capítulo anterior, es la función evolutiva que cumplen nuestras emociones la que explica qué tipo de objetos generan qué reacciones emocionales. Nuestras emociones tienen la función de ayudar a nuestra supervivencia (en sentido darwinista) y por tanto ellas se disparan cuando necesitamos reaccionar. Dijimos también que introducir la teoría de Darwin le permite a James hacer recaer la noción de función en la ciencia natural, en lugar de hacerla depender del plan divino como en Descartes. Esto será vital para la naturalización que ofrece Prinz.

Como puede notarse, James apenas esboza el programa de naturalización, aunque aporta elementos importantes para dicho proyecto. James nota que la psicología parece requerir una explicación que se valga tanto de la fisiología como de la descripción de nuestros estados mentales, y que correlacione ambos eventos. En este sentido, James busca una naturalización de la mente. La introducción de la noción de función ayuda para este propósito, pues permite ver en las emociones funciones naturales y explicar —en el domino de la ciencia natural— cómo ellas se disparan frente a ciertos objetos. En otras palabras, la noción de función es un primer paso para la naturalización. No obstante, James todavía no dispone de una teoría clara de la mente que permita dar cuenta de manera precisa de la intencionalidad de nuestra mente. Esta es precisamente una de las tareas que —según James— debe cumplir la filosofía para que la

psicología pueda convertirse en ciencia natural (1892/2001, p. 331). Es aquí donde Prinz, de la mano de Dretske, toma esta tarea e introduce una naturalización más poderosa en el programa de naturalización de las emociones.

Prinz, Dretske y la naturalización de la intencionalidad. Hemos dicho que Prinz introduce la teoría de la representación mental de Dretske para naturalizar la intencionalidad de las emociones. Además, hemos expuesto algunos de los aspectos de esta teoría, como la manera en que se explica la formación de conceptos y algunos rasgos de las emociones. Veamos ahora cómo se explica la intencionalidad, acudiendo a la versión presentada por Dretske. Como intentaré mostrar, esta teoría se vale de elementos que ya estaban presentes en las teorías de Descartes y de James, a saber, las nociones de representación y función.

En primer lugar, Dretske suscribe la tesis representacional, según la cual "todos los hechos mentales [son] hechos representacionales, [...]. Mi experiencia de un objeto es la totalidad de las formas en las que el objeto me aparece, y la manera en que el objeto me aparece es la manera en que mis sentidos lo representan" (Dretske, 1995, p. 1).[45] Esto significa que mi experiencia de un objeto no es más que la manera en que mis sentidos lo presentan ante mí y a esta presentación-ante-mí la llamamos una "representación".

[45] Traducción libre del autor, el texto original dice: "[…] all mental facts [are] representational facts […]. My experience of an object is the totality of ways that object appears to me, and the way an object appears to me is the way my senses represent it".

Esta representación, como sucedía en Descartes, no tiene por qué hacer aparecer al objeto con todas sus propiedades. Un sistema representacional, para Dretske, solo representa una o algunas propiedades de un objeto. Más aún, un sistema representa únicamente aquellas propiedades para las que el sistema fue diseñado para representar. En palabras del autor: "un sistema, S, representa una propiedad, F, si y sólo si S tiene la función de indicar (proveer información de) la F de un cierto dominio de objetos" (1995, p. 2).[46] El ejemplo que ofrece Dretske es el de un velocímetro. El velocímetro representa aquello para lo que fue diseñado, aquello que es su función representar, i.e., la velocidad de un vehículo. No podría representar, por ejemplo, el color del vehículo, toda vez que esta no es su función.

Puede notarse de inmediato que este ejemplo apela a una máquina diseñada por el ser humano. Por esto, Dretske distingue entre sistemas convencionales y naturales. Los sistemas convencionales son aquellos cuyas funciones son, justamente, convencionales, como el velocímetro. Este representa en virtud de nuestras convenciones. Como puede esperarse, los sistemas naturales son aquellos que representan en virtud de alguna función natural. ¿Cómo entender estas funciones naturales? En este caso, son funciones biológicas de un organismo (Dretske, 1995, p. 8). Nuestros sistemas sensoriales, por ejemplo, representan los colores o

[46] Traducción libre del autor, el texto original dice: "[...] a system, S, represents a property, F, if and only if S has the function of indicating (providing information about) the F of a certain domain of objects".

texturas de los objetos porque tienen la función biológica de hacerlo. Como podemos anticipar, esta noción de «función biológica» puede recaer en una teoría como la darwinista, en la cual un organismo adquiere sistemas de representación porque ellos cumplen una función en su supervivencia.

Ahora bien, ¿qué determina que un sistema represente un objeto determinado? En otras palabras, ¿qué explica la intencionalidad de un sistema representacional como la mente? Primero que todo, hemos de notar, como bien señala Dretske, que el que un sistema represente un objeto como que tiene determinadas propiedades no implica que el objeto las tenga. Retomando el ejemplo de Putnam, podemos hacer una caricatura de Winston Churchill con una nariz enorme; si bien la caricatura representa a Churchill teniendo la propiedad de tener una nariz grande, de allí no se sigue que Winston Churchill tenga en efecto dicha propiedad. Esto hace eco de la idea cartesiana de que nuestras representaciones no tienen por qué parecerse a aquello que representan. Entonces, si no es por las propiedades representadas, ¿cómo fijamos el objeto de nuestra representación?

Es aquí donde se ofrece una explicación naturalizada de la intencionalidad. Para Dretske, un sistema tiene un objeto si cumple dos condiciones. La primera de estas condiciones consiste en que el sistema tenga la función de representar la propiedad F de ese objeto, como dijimos anteriormente. El velocímetro representa la velocidad de mi vehículo porque tiene esta función. La segunda condición estriba en que el sistema esté en una relación C con el objeto a representar. Esta relación C es de índole contextual, es decir, es relación

entre el sistema y el contexto en el que se encuentra. Decimos, por ejemplo, que el velocímetro representa la velocidad de mi vehículo porque está en cierta relación con mi vehículo, y no con otros; es decir, el contexto en el que está el velocímetro es el de estar conectado a mi vehículo y no a otro. En ese orden de ideas, podemos decir que un sistema S representa una propiedad F de un objeto O si y solo si *(1)* S está en la relación contextual C con O y *(2)* si S tiene la función de representar F de O.

Nótese que esta explicación de la intencionalidad de la representación se vale únicamente de los términos de 'función' y 'relación contextual'. Ambos términos pueden ser extraídos de la ciencia natural para nuestros estados mentales. La noción de función es extraída aquí de la biología; en particular, la teoría de Darwin juega un papel esencial. La noción de 'relación contextual' puede interpretarse, de manera laxa, como una relación causal. Esta es, por ejemplo, la idea de Prinz (2004b, p. 53). Luego, vemos que aquí se ofrece una explicación de la intencionalidad en términos naturalizados.

Pese a las posibles objeciones que puedan ofrecerse a este intento de naturalización, quiero resaltar el tipo de teoría que se requiere para dar cuenta de la intencionalidad de las emociones. El aporte de Prinz en esta dirección es el de formular una teoría de las emociones basada en esta teoría de la representación. En términos de Dretske, para Prinz las emociones son sistemas representacionales que tienen la función de representar alguna propiedad de los objetos de nuestras emociones. Estas propiedades son los temas

relacionales básicos: una emoción representa una relación de un objeto con nuestro bienestar. Así, la intencionalidad de las emociones se explica de la misma manera en que se explican otras representaciones mentales, esto es, acudiendo a su función y su relación contextual (situacional). Por ejemplo, sentir alegría por el ascenso en mi trabajo se explica porque mi ascenso está en la relación contextual requerida para que se dispare mi sistema de felicidad. Esta relación contextual es definida como un tema relacional básico: mi ascenso implica un progreso hacia mis metas.

¿Dónde entran entonces las reacciones corporales? Así como el velocímetro representa la velocidad contando (rastreando) las rotaciones del eje del vehículo, Prinz sostiene que las emociones representan temas relacionales básicos rastreando los cambios que ocurren en el cuerpo. En este sentido, podemos sintetizar la teoría de Prinz de la siguiente manera:

> Yo suscribo que las emociones rastrean temas relacionales básicos registrando cambios en el cuerpo. [...] Tal y como los conceptos de perros rastrean a los perros vía su pelo, el temor rastrea peligros vía palpitaciones del corazón. Las emociones son corpóreas. Ellas representan temas relacionales básicos, pero lo hacen percibiendo cambios corporales (Prinz, 2004b, p. 68).[47]

[47] Traducción libre, el texto original dice: "I submit that emotions track core relational themes by registering changes in the body. [...] Just as concepts of dogs track dogs via furriness, fears track dangers via heart palpitations. Emo-

Nótese entonces cómo la teoría de Dretske se vale de las nociones de representación y función. Para el caso de las emociones, vimos que Descartes tiene observaciones que apuntan hacia la primera, la noción de representación, mientras que James se acerca a la noción naturalizada de función. Prinz, entonces, logra insuflar una teoría que recoge ambas nociones e intenta explicar con ellas la intencionalidad de las emociones. Así, se intenta dar cabida a la intencionalidad en el mundo natural. De ahí que podamos decir que el programa de naturalización de las emociones requiere —y obtiene con Prinz— una naturalización de la intencionalidad.

Siguiendo la matriz de Lakatos, vemos en este caso una instancia de progreso teórico paulatino en el programa de naturalización de las emociones. Partimos, primero, de la formulación de un núcleo central (la definición de las emociones como la percepción de reacciones corporales) y vimos cómo esta nos hace tropezar un el problema de la intencionalidad. Sobre este problema vemos a su vez movimientos en las hipótesis auxiliares alrededor del programa, en los que encontramos distintas aproximaciones al problema hasta la introducción de una teoría de la intencionalidad. Con ello podría decirse que el programa adquiere poder predictivo pues ahora podemos explicar y predecir casos como el mencionado temor que sentimos por una serpiente.

---

tions are embodied. They represent core relational themes, but they do so by perceiving bodily changes".

Lo que viene a continuación es una problematización de este proyecto. Como podremos ver, varias de las observaciones de Wittgenstein parecen mostrar algo extraño sobre el proyecto de naturalización. Si tenemos éxito, veremos que el problema de la naturalización de la intencionalidad como problema filosófico, en el fondo, carece de sentido y, si ese es el caso, esta crítica puede expandirse a la teoría naturalista sobre las emociones.

3. Wittgenstein: de la ciencia a la gramática

Hasta ahora hemos expuesto el programa de naturalización de las emociones, y hemos señalado como uno de sus problemas el de la naturalización de la intencionalidad. Permítaseme ahora alejarnos por un momento del tema de las emociones y de su intencionalidad y concentrémonos en algunas de las anotaciones de Wittgenstein a propósito de temas relacionados con la psicología y la gramática de los vocablos mentales. Si bien intentaremos aplicar estas observaciones, primero debemos tener algo de claridad sobre cómo procede la filosofía practicada a la manera wittgensteiniana, esto es, qué significa la filosofía como terapia. Esto nos permitirá adoptar una perspectiva diferente sobre la investigación acerca de las emociones, de modo que podamos pasar de un programa de investigación científica a una elucidación gramatical. Así, haremos primero un breve sondeo del carácter terapéutico de la filosofía de Wittgenstein. Luego haremos un despliegue de algunas de las herramientas que nos ofrece el autor para estudiar conceptos psicológicos. Con ello intentaremos en el siguiente capítulo aplicar estas herramientas a nuestro caso de estudio, el programa de naturalización de las emociones.

3.1. La metodología: la filosofía como terapia

Para comprender la metodología de Wittgenstein para tratar problemas filosóficos, debemos primero entender cómo es que, a grandes rasgos, la filosofía llega a sus problemas y sus métodos de solución. En otras palabras, debemos ver cuál es la pretensión de la filosofía tradicional y cómo Wittgenstein supone un giro en la manera de tratar los problemas filosóficos.

De manera general, un filósofo tradicional identifica un problema, hace un recuento de otros pensadores que se han ocupado de él y procede a decir por qué las ideas y aproximaciones de sus predecesores han sido equivocadas. Tras identificar el error de los anteriores, el filósofo da pie a una elaborada teoría que, en su opinión, soluciona el problema planteado (Hacker, 1972/1986). Nótese que este es el proceder tanto de Descartes, como de James, como de Prinz. Los tres comienzan sus textos aludiendo a los errores que los demás han cometido al tratar las emociones. Descartes, por su parte, comienza sosteniendo que "No hay nada en que aparezca mejor cuán defectuosas son las ciencias que tenemos de los antiguos que en lo que han escrito de las pasiones" (AT, XI, 327). James afirma que

> Los fisiólogos que, durante algunos años anteriores, han estado explorando industriosamente las funciones del cerebro, han limitado sus intentos de explicación a sus actividades cognitivas y volitivas. [...] Pero la esfera *estética* de la mente, sus añoranzas, sus placeres y sus

> dolores, y sus emociones, ha sido tan ignorada en todas sus investigaciones [...] (James, 1884, p. 188).[48]

Finalmente, Prinz dice: "En los capítulos que siguen, arguyo que los principales intentos de dirigirse a estos problemas interconectados son inadecuados. Pero mi objetivo primerio es constructivo, no crítico. Propondré una teoría positiva de las emociones" (Prinz, 2004b, p. 19). Vemos, así, que todos ellos comienzan aduciendo que ningún otro ha tratado satisfactoriamente el problema del que ellos se ocuparán y del que sí ofrecerán una teoría apropiada.

Este proceder en filosofía, sin embargo, supone ya que los problemas filosóficos exigen una solución, esto es, que son problemas genuinos. Aquí se presupone que el filósofo tiene un problema legítimo y un método de solución. Wittgenstein se separa de forma radical de esta manera de proceder, lo que determina un rumbo completamente diferente de pensamiento. La filosofía, en este sentido tradicional, supone que su labor es la de corregir la forma en la que entendemos las cosas. Cuando un filósofo se pregunta "¿Qué es el conocimiento?", quiere ofrecernos criterios para usar correctamente el término 'conocer'. Lo que el filósofo no ve —piensa Wittgenstein— es que no necesitamos tal corrección. El lenguaje ordinario, por sí solo, no tiene

---

[48] Traducción del autor del texto original: "The physiologists who, during the past few years, have been so industriously exploring the functions of the brain, have limited their attempts at explanation to its cognitive and volitional performances. [...] But the aesthetic sphere of the mind, its longings, its pleasures and pains, and its emotions, have been so ignored in all these researches [...]".

ningún error en el sentido en el que lo espera el filósofo —y, si lo tiene, son las prácticas mismas las que nos indican cómo proceder—, el error es, más bien, pensar que el lenguaje ordinario no basta y que necesitamos un lenguaje más refinado y depurado para entender nuestros conceptos (por ejemplo, el lenguaje de la lógica) (IF §98). En otras palabras, la ilusión es pensar que necesitamos teorías que nos enseñen a usar palabras que ya sabemos usar.

Para no caer en esta tentación, Wittgenstein propone una nueva metodología que, en lugar de generar una teoría filosófica, busca mostrarnos que los problemas filosóficos en relación a la esencia o fundamento del lenguaje y de sus conceptos en realidad no están allí. En otras palabras, la metodología de Wittgenstein busca liberarnos de la ilusión de que hay algún problema filosófico profundo, y de que necesitamos un lenguaje depurado para entender nuestros conceptos. Lo que debemos hacer, dice Wittgenstein, es atender a nuestros usos comunes, a los modos de uso (cuan diversos puedan ser) de nuestro lenguaje.

Ver estos usos, sus relaciones y cómo estos se conectan íntimamente con nuestras formas de vida —nuestras convenciones— implica lograr lo que el autor llama una "visión sinóptica" (IF §122). Debemos ver en nuestro lenguaje una gran cantidad de conexiones entre distintos usos y nuestras prácticas lingüísticas, sin pensar que debe haber una gran esencia última subyacente a todos los usos de una palabra. Así, podremos ver que los problemas que nos aquejan como filósofos no son más que fórmulas gramaticales confusas, o traslapos entre distintos usos de un mismo término, u

otro sinfín de posibles tropiezos que el lenguaje nos lleva a cometer.

Sin embargo, esta visión sinóptica no es simplemente una visión general. Con ella no queremos llegar a generalidades sobre los conceptos, pues ello sería precisamente alcanzar la esencia oculta mediante una generalización. Lo que queremos, al contrario, es reconocer la variedad, las diferencias y las semejanzas, sin intentar dar con una teoría homogénea sobre los usos de los conceptos. Cuando procedamos en el mar de las confusiones filosóficas, intentaremos estudiar casos, ejemplos, que nos muestren las distintas formas en las que se usan los conceptos, sus distintas relaciones en distintos contextos, de modo que logremos apreciar la amalgama de usos posibles. Eso se hace mediante el uso de los "juegos de lenguaje", herramientas de contraste para examinar estos casos.[49] Al revisar diferentes juegos de lenguaje tenemos la expectativa de ver en los conceptos las múltiples relaciones posibles, con la esperanza de que ellas nos aporten una perspectiva diferente sobre nuestros conceptos. Con esta nueva perspectiva —la visión sinóptica— esperamos desprendernos de nuestra confusión. En este sentido, la visión sinóptica se adquiere, no al ver desde una generalidad, sino al ir a casos concretos del lenguaje. Así, el estudio del wittgensteiniano parte de casos concretos en

[49] Como es común en la filosofía de Wittgenstein, la noción de 'juego de lenguaje' tiene una amplia polisemia, es decir, tiene distintas acepciones a lo largo de la obra. De manera general, podemos entender esta noción, como hemos dicho, como una herramienta de contraste que nos permitiría apreciar distinciones y semejanzas entre distintos usos del lenguaje.

los que usamos las palabras ordinariamente, y poco a poco, caso a caso, va percatándose de que lo que hay en el fondo es una variedad de usos y no un problema filosófico real.

Ello, no obstante, no debe confundirse con una etnografía del lenguaje. Es posible que estos casos sean ficticios. Más aún, a menudo Wittgenstein inventa los casos a estudiar a lo largo de su obra. Lo importante aquí es que los casos, sean ficticios o reales, nos permitan ver con claridad que no hay problema profundo, que nuestros usos ya muestran la manera en la que entendemos los conceptos, y que no es necesaria una intrincada teoría filosófica para comprenderlos. En este sentido, la descripción que estos casos deben ofrecer no es una descripción empírica etnográfica, sino una descripción de las posibilidades lógicas —o, mejor, gramaticales— del uso de un concepto.[50]

Uno de los casos paradigmáticos de aplicación de esta metodología es el que Wittgenstein estudia alrededor de la palabra 'juego'. Si preguntamos qué es un juego, podemos estar tentados —como filósofos— a buscar la esencia subyacente a todos los juegos. No obstante, Wittgenstein

---

[50] Pareciera que Wittgenstein es un filósofo del sentido común. Sin embargo —como señala Lazerowitz— Wittgenstein es distinto de otros filósofos del sentido común como Moore —y, podemos aventurarnos a decir, alguien como Hume—. Para Moore, el sentido común presenta la solución del problema filosófico. El filósofo [tradicional] parece no ver el sentido común; lo que hay que hacer, luego, es repetirle al filósofo cómo pensamos ordinariamente. Si bien puede sonar parecido a Wittgenstein, esto supone justo que el filósofo es ciego al sentido común. No obstante, como señala el mismo Wittgenstein, no es que el filósofo sea ciego, sino que se confunde en medio del uso del sentido común. Véase Lazerowitz (1968) y CA (p. 95).

nos invita es a ver que no hay tal esencia. Si pensamos en juegos concretos, veremos que hay juegos de todo tipo (cartas, tablero, pelota, etc.), con reglas definidas o sin ellas, con ganadores y perdedores o sin ellos, etc. Este hecho, la variedad entre los distintos juegos, no responde a ninguna causa particular, se trata simplemente de una forma en la que hemos aprendido a vivir. En otras palabras, si alguien nos pregunta por qué los juegos son tan distintos y aun así los llamamos con la misma palabra, tendremos que decir: "porque así hablamos". Llegados a este punto, cualquier pregunta posterior carece de sentido, pues hemos llegado a una base que, si bien no configura un fundamento último o profundo de nuestro concepto de juego, es un hecho primitivo de nuestras prácticas. Este punto, en el que ya no tiene sentido una pregunta posterior, en donde debemos decir "así somos", es lo que Wittgenstein llama un *protofenómeno*. Aquí la descripción de las distintas maneras de jugar nos lleva a ver que no hay nada común o esencial a un juego y que el concepto remite es a un conjunto de prácticas que de una u otra forma aprendimos a llamar así.

Notemos en esa manera de proceder que no hemos ofrecido ninguna hipótesis sobre qué podría ser común a los juegos, o qué podría fundamentar la variedad que hemos detectado. En este sentido, no nos interesa *explicar* la variedad, sino solo describirla. Hemos visto de manera amplia los distintos juegos, y al tener esa perspectiva, esa visión sinóptica, hemos advertido que no hay nada que deba ser explicado. Si la terapia funciona, ya no debemos

atormentarnos por la esencia de los juegos, pues habremos aprendido a ver que en el fondo no hay más que convención.

Si nos preguntamos, por ejemplo, por qué decimos que el metro patrón de París mide un metro (IF §50), debemos reconocer en el fondo que esto no es más que una convención en la que una vara modela el significado de "medir un metro". Si al medir una mesa esta mide un metro y medio, lo que estamos diciendo es que puedo hacer caber una vara y media en su longitud. A continuación, podemos preguntarnos: ¿y por qué esa vara que está en París es la que modela el significado de "medir un metro"? ¿Qué tiene esa vara de especial? Podría, por supuesto, conducir muchos experimentos alrededor de la vara, intentando encontrar una propiedad oculta especial sobre lo que significa "medir un metro". No obstante, la invitación de Wittgenstein frente a estos casos es responder: no hay nada especial, nada profundo. En el fondo, solo hay convención. Convenimos en algún momento que la vara iba a ser el paradigma de longitud, y eso es todo lo que podemos decir. En este momento, la querella filosófica (o científica, si es el caso) por la naturaleza de nuestras mediciones de longitud debe cesar, pues hemos encontrado que no hay nada profundo que determine que las cosas midan un metro. Aquí debemos parar la pregunta filosófica. A esto es a lo que Wittgenstein llama "llevar a la filosofía al reposo" (IF §133).

Así las cosas, lo que se intentará practicar es una suerte de investigación en la que no se busca la esencia del lenguaje, sino que se intenta ver en el lenguaje una variedad de usos que, en el fondo, no esconden más que lo que Wittgenstein

llama "protofenómenos". En palabras del autor: "Nuestro error es buscar una explicación allí donde deberíamos ver los hechos como 'protofenómenos'. Es decir, donde deberíamos decir: *éste es el juego de lenguaje que se está jugando*" (IF §654; énfasis en el original). En el fondo, lo que subyace al lenguaje no es más que la convención, esto es, las maneras en las que hemos aprendido a usar los conceptos en determinadas prácticas lingüísticas.

Notemos que en el parágrafo citado anteriormente Wittgenstein usa el término 'explicación' para referirse al error. Podríamos resumir esta afirmación diciendo que nuestro error está en buscar explicar algo que no necesita ser explicado. ¿Qué significa esto? Explicar el lenguaje, por ejemplo, significaría buscar una teoría general que nos ofrezca la esencia del lenguaje. Esta es la teoría que se busca en el *Tractatus* y en las obras de los autores anteriormente citados. Sin embargo, hemos dicho ya que la imagen que quiere desarrollar Wittgenstein intenta no ver en el lenguaje una esencia oculta a develar (por ejemplo, mediante análisis), sino una variedad de usos. Si este es el caso, entonces lo que estamos buscando no es una explicación; estamos buscando una "descripción" de esos distintos usos. En palabras de Wittgenstein:

> No puede haber nada hipotético en nuestras consideraciones. Toda *explicación* tiene que desaparecer y sólo la descripción ha de ocupar su lugar. Y esta descripción recibe su luz, esto es, su finalidad, de los problemas filosóficos. Éstos no son ciertamente empíricos, sino que

> se resuelven mediante una cala en el funcionamiento de nuestro lenguaje, y justamente de manera que éste se reconozca: *a pesar de* una inclinación a malentenderlo. Los problemas se resuelven no aduciendo nueva experiencia, sino compilando lo ya conocido. La filosofía es una lucha contra el embrujo de nuestro entendimiento por medio de nuestro lenguaje (IF §109; énfasis en el original).

En el caso del metro patrón de París hemos encontrado una convención en el fondo de nuestras prácticas de medición métrica. Lo que hemos hecho es, entonces, *describir* la manera en la que hemos modelado el uso de un concepto, esto sin haber ofrecido una explicación sobre la naturaleza profunda de la medición. Es a través de la descripción de este tipo de usos y convenciones que debemos acallar nuestra confusión filosófica y detener una investigación por los fundamentos de los conceptos.

En resumen, la investigación filosófica wittgensteiniana no apunta a una revelación del fundamento del lenguaje. No estamos buscando una teoría que nos revele la naturaleza profunda de nuestros conceptos, ni un método unívoco que nos revele algo que nos era oculto. Debemos, precisamente, ver mediante el lenguaje, mediante sus usos y prácticas, que no hay nada profundo allí. Todo esto nos debe mostrar, al final, que la investigación filosófica no debe ir tras la quimera de la esencia misma de los conceptos. Así, lo que hay en el fondo no es una esencia, sino una forma de actuar (Wittgenstein, 1969/2006, §402), una convención

sobre la manera en que practicamos el lenguaje. La filosofía tradicional, aquella que construye elaboradas teorías sobre cómo debemos usar el lenguaje y sobre la esencia misma de los conceptos, es producto de numerosas confusiones en las que nos sumergimos gracias a las trampas del lenguaje. De ahí que la forma en que se practica la filosofía wittgensteiniana sea tan diferente, incluso, él mismo llegó a llamarla "uno de los descendientes de la materia que solía llamarse 'filosofía'" (CA, p. 59) para diferenciarla de la filosofía tradicional.

De este tipo de consideraciones, y para introducir el segundo punto que quiero resaltar, algunos han sugerido que Wittgenstein despliega una nueva *teoría* sobre el lenguaje, una teoría que identifica el significado con el uso de las palabras y no con la referencia y con la cual podemos resolver toda clase de problemas filosóficos. No obstante, es central recordar que para Wittgenstein la filosofía no se encarga de construir teorías.

Ahora bien, en el ámbito de los conceptos psicológicos, la aplicación de las terapias wittgensteinianas tiene ciertas dificultades que es importante advertir con antelación. No son pocas las confusiones que surgen al hablar de lo mental y a menudo estas se traslapan y entremezclan para crear confusiones aún mayores. Intentaremos, entonces, atender a algunas de las recomendaciones de Wittgenstein para estudiar estos conceptos, procurando siempre mantenernos en el plano de la descripción de los usos, esto es, el plano de la descripción gramatical.

3.2. Aclaraciones sobre los conceptos psicológicos

Sobre la tarea de una psicología científica. Como hemos adelantado, hay varias dificultades que atraviesan la investigación sobre los conceptos psicológicos. La primera que me gustaría resaltar es la aparente necesidad de una teoría sobre lo mental. Cuando hacemos una teoría de la mente queremos que la teoría nos explique qué sucede en nuestra vida mental, qué significa decir que creemos, que tenemos intenciones o que pensamos algo. Cuando un filósofo como Descartes, James o Prinz habla de las emociones, afirma que está buscando una definición verdadera sobre el fenómeno que está estudiando. Su meta será ofrecernos una explicación sobre la naturaleza oculta de su fenómeno. Quiere enseñarnos algo que no sabíamos antes. En el caso de los tres filósofos, ellos quieren enseñarnos que las emociones son esencialmente reacciones corporales. Para apoyar su querella conceptual, para definir el concepto de emoción, aluden a la evidencia empírica, y sostienen que su teoría es correcta porque la experiencia lo demuestra. Si en algún punto hay dificultades, estas son producto de dificultades generales de la psicología y no de su teoría.

En esto, Prinz es directo de manera especial. El autor constantemente acude a la evidencia empírica, incluso prefiriéndola frente al análisis conceptual (2004b, p. 29). Al discutir las dificultades de distinguir las emociones básicas (aquellas que no son reducibles a otras emociones) de las emociones complejas, dice:

> Creo que la carencia de consenso no es una causa para el escepticismo. [...] Realmente, [de ser así] toda teoría científica sería vulnerable. El desacuerdo es rampante en ciencia. El desacuerdo muestra que alguien debe estar equivocado, pero no muestra que todos estén equivocados. La diversidad de listas de emociones básicas no arroja más duda sobre el programa de emociones básicas que la que la teoría de los elementos de Aristóteles arroja sobre la tabla periódica moderna. La ciencia de la emoción *está en su infancia,* y las listas actuales reflejan especulaciones tempranas (Prinz, 2004b, p. 91; énfasis fuera de texto).

Su idea, no solo en cuanto a las emociones básicas sino en cuanto a todo el programa de investigación sobre las emociones, es recopilar más evidencia que nos muestre su naturaleza.

Sin embargo, Wittgenstein es especialmente crítico sobre esta manera de proceder. Para él, la evidencia empírica no puede ayudarnos en la elucidación de los conceptos de la psicología. Citando al autor:

> La confusión y esterilidad de la psicología no se puede explicar por el hecho de que es una "ciencia joven"; no se puede comparar su estado, por ejemplo, con el de la física en sus comienzos. (En todo caso más bien con el de ciertas ramas de la matemática. Teoría de conjuntos). En efecto, en psicología existen métodos experimentales y *confusión conceptual*. (Así como en el otro caso

mencionado existen confusión conceptual y métodos de demostración).

> La presencia del método experimental nos hace creer que ya disponemos de los medios para librarnos de los problemas que nos inquietan; cuando en realidad problemas y métodos pasan de largo sin encontrarse (IF, II §XIV, énfasis en el original; OFP I §1093).

La psicología no es simplemente una ciencia joven, en el sentido en que no es un programa de investigación constituido a espera de ser desarrollado. En ella todavía encontramos —para Wittgenstein— numerosas confusiones conceptuales. A diferencia de los problemas empíricos, los problemas conceptuales no cuentan de entrada con un método de solución. Para dar un ejemplo, cuando un físico intenta resolver un problema de su ciencia ya tiene a su mano diversas herramientas matemáticas y experimentales para resolverlo. El físico, a continuación, hará uso de ellas para salir de su problema. En filosofía, en cambio, si lo que queremos es ocuparnos de problemas conceptuales, debemos reconocer que no hay un único conjunto de métodos y reglas de solución; en ese sentido, Wittgenstein sostiene que “La forma de un problema filosófico es: no sé salir del atolladero” (IF §123). En otras palabras, no está dada la manera de solucionar un problema filosófico, a diferencia de como ocurre en una ciencia. Para atacar un problema conceptual no tenemos una vía clara para proceder, no sabemos bien cómo alcanzar ni qué sería exactamente una solución.

Wittgenstein constantemente nos reitera que lo que nos ocupa en el caso de la psicología no es una teoría científica, esto es, no nos ocupa un problema empírico, sino un problema conceptual. En este sentido, no nos ocupa un problema que podamos solucionar mediante la experimentación, como es el caso del físico. Para salir de los problemas conceptuales, no elaboramos teorías ni las confirmamos en el mundo. La propuesta de Wittgenstein, recordemos, es tener una visión sinóptica del lenguaje, i.e., una descripción de la variedad de usos y las relaciones entre distintas formas del lenguaje. En el fondo, no nos interesa explicar —y del mismo modo tampoco nos interesa experimentar— sino que nos interesa describir.

Cuando pasamos por alto la importancia de esta descripción e intentamos ofrecer una teoría filosófica, queremos construir un nuevo lenguaje que nos ayude a definir con claridad el concepto del que nos ocupamos. En el caso del programa de naturalización de las emociones queremos construir un lenguaje fisiológico que nos ayude a entender con claridad qué es una emoción. Sin embargo, el lenguaje ordinario, para Wittgenstein, está bien como está; lo que necesitamos, en lugar de construir una "nueva notación" (CA, p. 93), es describir los usos, cuan variados sean, de las palabras. En otras palabras, no necesitamos un nuevo lenguaje, sino que necesitamos hacer caso de lo que ya tenemos ante nosotros en el lenguaje ordinario.

Esta descripción de los usos de los conceptos, aplicada a los conceptos psicológicos, es especialmente compleja, toda vez que pareciera que la psicología quiere hacer usos

técnicos de términos que usamos en nuestro día a día. En otras palabras, cuando describimos conceptos sobre nuestra vida mental (pensamiento, sensación, emoción), es muy fácil caer en confusiones, pues tenemos la tentación de creer que la psicología —en tanto ciencia de lo mental— nos dirá con más claridad y precisión cuál es la naturaleza de estos conceptos. Baker y Hacker (1982) lo resumen de la siguiente manera:

> Los conceptos que la filosofía de la mente [y la filosofía de la psicología] deben clarificar son conceptos cotidianos. Ellos no son términos técnicos del arte de una ciencia avanzada. Cuando un filósofo intenta elucidar el concepto de mente, de pensar o de creer, de desear o de querer, de sensación o de emoción, él está en el trabajo de clarificar conceptos comunes que todos empleamos en nuestro discurso diario. Es también notable que los conceptos básicos de la psicología empírica (no como los conceptos fundamentales de la física) son los conceptos ordinarios de la vida diaria (p. 228).[51]

---

[51] Traducción libre del autor del original: "The concepts which philosophy of mind must clarify are everyday concepts. They are not the technical terms of art of an advanced science. When a philosopher attempts to elucidate the concept of mind, of thinking or believing, of wanting or intending, of sensation or emotion, he is in the business of clarifying common-or-garden concepts which we all employ in our daily discourse. It is also noteworthy that the basic concepts in empirical psychology (unlike the fundamental concepts of physics) are the ordinary concepts of daily life".

Cuando un científico nos quiere hablar de las emociones pareciera querer corregir usos ordinarios por usos más precisos que él ha descubierto apoyado en experimentos. En esta línea de razonamiento, filósofos como James o Prinz creen que una mayor evidencia puede develarnos los aspectos ocultos de nuestras emociones. Al hacer ciencia, queremos aprender qué significa realmente la palabra 'emoción'. Podemos decir, en este punto, que para la ciencia el concepto de emoción es en cierto modo desconocido y que su labor entonces es aclararlo mediante investigación empírica, i.e., mediante la aplicación de una teoría y la construcción de hipótesis.

¿Pero realmente es desconocido el concepto de 'emoción'? ¿Es aclarar el concepto de 'emoción' una tarea para una investigación empírica? Pensemos en otro concepto científico, digamos, el concepto de 'electrón'. En principio, este concepto tiene sentido en el marco de una teoría física que lo defina. En este orden de ideas, necesitamos de la ciencia para que ella nos enseñe qué es un electrón. Sin embargo, si bien necesitamos de una ciencia que nos enseñe el significado de la palabra 'electrón', nosotros ya conocemos de antemano el uso de las palabras que estudia la psicología. Cuando me preguntan "¿Qué estás pensando?" o "¿Dónde te duele?", puedo responder sin mayores tropiezos: "Pienso en lo que aprendí hoy en clase" o "Me duele el brazo". Como podemos notar, usamos conceptos psicológicos en nuestra vida cotidiana, a diferencia de conceptos técnicos como "carga de un electrón". Estos últimos requieren del contexto de una teoría que los aclare; en cambio, usamos

conceptos psicológicos sin dificultad incluso antes de hacer teorías sobre ellos.

¿Por qué, entonces. nos parece extraño el concepto de "emoción" cuando lo analizamos científicamente? ¿De dónde proviene esta ilusión de que necesitamos de una teoría que nos enseñe su uso correcto? La meta del análisis wittgensteiniano es ver qué nos parece extraño, no porque oculte algo que la ciencia se encargará de develar, sino porque estamos haciendo malos usos del lenguaje. Podemos decir, en este punto, que uno de estos malos usos radica en que confundimos el papel de la investigación empírica con el de la elucidación conceptual. El filósofo como Descartes, James o Prinz busca crear una teoría empírica sobre las emociones y este trabajo debe enseñarnos qué es —en nuestro caso— una emoción. No obstante, como sostiene Hacker (2004/2009).

> La filosofía no puede hacer descubrimientos empíricos, y no es su tarea producir teorías empíricas. Su tarea es describir las estructuras conceptuales en términos en los que articulamos nuestra experiencia y sus objetos. Entonces, mientras ella no debería proponer *teorías empíricas* de las emociones, ella puede elucidar el marco *conceptual* para la investigación experimental. Por lo tanto, ella también puede prevenir confusiones conceptuales que vician el diseño y las conclusiones inferidas de los experimentos (p. 43, énfasis en el original).[52]

[52] Traducción libre del autor del original: "Philosophy cannot make empirical discoveries, and it is not its task to produce empirical theories. Its task

En este orden de ideas, nuestra investigación sobre las emociones, desde el punto de vista wittgensteiniano, no debe pretender un hallazgo empírico. En otras palabras, nuestra investigación conceptual no debe confundirse con la investigación empírica. La primera es, si se quiere, lógicamente anterior a la segunda. El carácter juvenil de la psicología, como lo sugiere Wittgenstein, no se debe a una falta de experimentación, sino a que no hemos logrado el primer paso, no tenemos claridad sobre los conceptos que tratamos. Pero esta falta de claridad, como hemos dicho también, no proviene de la falta de investigación, proviene de nuestros malos usos.

Wittgenstein identifica una gran variedad de dificultades en el uso de los conceptos psicológicos cuando se trata de la investigación científica. Por un lado tenemos la ya mencionada confusión entre investigación empírica e investigación conceptual. Veamos a continuación otras confusiones que nos nublan la vista frente a los conceptos psicológicos.

Razones y causas. Un científico en su laboratorio ve un fenómeno y se pregunta "¿Por qué ocurre esto?" ¿Qué respuesta quiere el científico? En una palabra, quiere una «causa». Él quiere saber qué acontecimiento antecedió a aquello que vio en el laboratorio, qué fue lo que generó el evento que acaba de ocurrir. Para llegar a la causa, el

---

is to describe the conceptual structures in terms of which we articulate our experience and its objects. So while it should not propose *empirical theories* of the emotions, it can elucidate the *conceptual* framework for experimental investigation. Hence too, it can prevent conceptual confusions that vitiate the design of, and the conclusions inferred from, experiments".

científico hace una conjetura, una *hipótesis* y la somete a prueba. Tras repetidos experimentos, el científico asegura que ha hallado la causa del fenómeno en cuestión. Tomemos ahora otro caso. Una persona llega llorando a nuestra casa y le preguntamos "¿Por qué lloras?" A continuación, esta persona nos responde "Porque mi gato ha muerto". Ella no conduce experimentos para mostrarnos que ha sido así, ni este enunciado nos da una hipótesis sobre la posible causa de su tristeza. Ella nos da una "razón".

Ambas preguntas, superficialmente, tienen la misma estructura: ¿Por qué X? No obstante, sus usos son distintos. Una de ellas pregunta por la causa y la otra, por la razón. Para Wittgenstein, es frecuente que esta forma gramatical compartida genere confusiones entre ambos tipos de respuesta (CA, p. 43). A veces respondemos con razones cuando se nos pregunta por una causa. Por ejemplo, si preguntamos "¿Por qué es usted tan agresivo?" buscando la causa, y se nos responde "Porque la gente es tonta" en lugar de algo como "Porque fui criado de tal manera". Por otra parte, a veces respondemos con causas cuando preguntamos por razones. Casos de estos vemos en el programa de naturalización de las emociones. Por ejemplo, si nos preguntamos "¿Por qué esta persona siente tristeza?" y respondemos "Porque los mecanismos de evolución han generado sistemas que se disparan frente a lo que es una pérdida y el gato representa una pérdida" en lugar de "Porque su gato ha muerto".

Para Wittgenstein, "De la causa puede decirse que uno no puede *conocerla*, sino sólo *conjeturarla*" (CA, p. 43, énfasis del original). Esto significa que, a diferencia de las

razones, las causas son hipótesis que podemos confirmar o rechazar según nuestros estándares empíricos y sobre las que podemos siempre tener dudas y demandas por una confirmación. Cuando decimos que un fenómeno causa otro, estamos haciendo una apuesta sobre el estado de cosas del mundo, apuesta que luego podemos abandonar. Las razones, en cambio, no son simples conjeturas. Cuando damos razones, por ejemplo de nuestra acción, no estamos diciendo "Quizás actué por esto y aquello, pero hace falta confirmación", sino que en cierto sentido sabemos bien cuál es nuestra razón. Para dar con nuestras razones, no hacemos experimentos ni elaboramos teorías. Podemos decir, en sentido laxo, que las razones hacen parte misma de la descripción de mi acción.[53]

Ahora bien, es importante aclarar que no debemos pensar que este "saber" que tenemos sobre nuestras razones sea gracias a un acceso privilegiado a nuestros estados mentales. No se trata aquí de que sepamos de nuestras razones por una suerte especial de introspección. Si fuera así, el estatus de nuestras razones sería similar al de una hipótesis, solo que en este caso la confirmación de esta hipótesis sería desde un punto de vista especial interno. Como veremos con más detalle más adelante, la idea de un acceso privilegiado a mis estados mentales es otra ilusión que debemos evitar. Para

[53] No deseo entrar aquí en discusiones sobre la filosofía de la acción. Si bien este punto tiene bastante discusión y una literatura vasta, no es el objeto del trabajo presente evaluar la naturaleza de nuestras descripciones de la acción. Por lo pronto, mi objeto es exponer la distinción que nos ofrece Wittgenstein a propósito de las razones y las causas.

Wittgenstein, no tiene sentido decir que no sabemos cosas sobre nuestros estados mentales; pero no porque tengamos un acceso privilegiado a ellos, sino porque enunciados como "Sé que tengo dolor de muelas" no designan hechos. Esto significa, en otras palabras, que enunciados como este describen formas de habla, formas de uso de un concepto, y no hechos en el mundo esperando ser develados. En el caso de "Sé que tengo dolor de muelas", el enunciado nos indica que sobre el dolor, en muchos casos, no tiene sentido decir "No sé si me duele mi muela o no". Por ejemplo, si vamos a un médico y él nos pregunta cómo sabemos que nos duele la muela, estaremos frente a una situación extraña en la que no sabemos qué contaría como procedimiento para "saber" tal cosa. Cuando vamos al médico con un dolor de muela, no aseguramos *saber* que tenemos dolor, simplemente decimos sentir dolor.

Lo mismo puede aplicarse ahora a las razones. Cuando decimos "Sé cuáles son mis razones", no estamos diciendo que hayamos encontrado una verdad profunda sobre nuestra vida mental. Lo que estamos diciendo con esta expresión es que —como con el dolor— no tiene sentido decir que no sabemos cuáles son nuestras razones. No se trata aquí de que tengamos un acceso epistémico privilegiado a ellas, como si pudiéramos observar de primera mano un hecho mental. Como dijimos, el enunciado aclara el uso de un concepto, el concepto de "razón". En ese sentido, decir que sabemos cuáles son nuestras razones es también una afirmación sobre una forma de habla, sobre una regla gramatical.

En contraste con ello, si bien pareciera que nuestra condición epistémica no nos dice nada nuevo sobre nuestras razones, sobre las causas yo sí puedo pedir confirmación, puedo tener dudas y demás. En este sentido, ofrecer una causa de un evento o de un fenómeno es formular una hipótesis. Para citar a Wittgenstein:

> La proposición de que una acción tiene tal y tal causa es una hipótesis. La hipótesis está bien fundada si se ha tenido un número de experiencias que, hablando toscamente, concuerden en mostrar que la acción es la secuela regular de ciertas condiciones, que entonces llamamos causas de la acción. Para conocer la razón que se tuvo para hacer un cierto enunciado, para actuar de un modo determinado, etc., no se necesita ningún número de experiencias acordes, y el enunciado de la razón no es una hipótesis (CA, p. 43).

Así las cosas, las causas son materia de investigación empírica, las razones, no. Para aclarar más esto, conviene comparar las razones con una demostración. En palabras del autor: "Dar una razón es como dar una operación de cálculo mediante la cual se ha llegado a un cierto resultado" (CA, p. 44). Las operaciones de un cálculo matemático no son propiamente causas del resultado, son, más bien, su justificación. Cuando damos una demostración de un cálculo, estamos justificando el porqué hemos llegado a tal resultado. Sobre estas demostraciones no hacemos ningún tipo de investigación empírica para ver si efectivamente

la demostración es la que usamos o no. Lo mismo ocurre con las razones, pues ellas no son materia de confirmación. Con ellas decimos el porqué hemos ejecutado una acción, el porqué hemos hecho una declaración, etc. En este mismo sentido, las razones son formas de justificación, más que conjeturas sobre lo que ha acontecido.

Al hacer investigación sobre los conceptos de la psicología, es frecuente encontrar debate sobre esta distinción. En un famoso artículo, por ejemplo, Davidson (1963/2002) sostiene que las razones son las causas de la acción. No obstante, si atendemos a las observaciones de Wittgenstein, debemos decir que en algunos casos las preguntas "¿Por qué…?" que recaen en las acciones preguntan por razones y en otras, por causas. Como hemos notado, la pregunta del científico "¿Por qué ha ocurrido tal cosa?" es una pregunta por la causa; sobre ella podemos hacer correr experimentos e investigaciones empíricas. En cambio, la pregunta "¿Por qué te sientes triste?", no busca la causa sino la razón; sobre ella no hacemos experimentos, pues con las razones ofrecemos una suerte de justificación más que una conjetura.

Como puede anticiparse (dada la última de estas preguntas), el análisis sobre razones y causas puede aplicarse al caso de las emociones. Cuando un fisiólogo pregunta "¿Por qué este paciente ha llegado a sentir tristeza?", busca una causa. El programa de naturalización de las emociones hace este tipo de preguntas. No obstante, cuando ordinariamente preguntamos "¿Por qué te sientes triste?", no buscamos que nos ofrezcan un listado de estados fisiológicos. Lo que buscamos es una razón. Sobre esto volveremos más adelante.

Es pertinente, primero, tener a la mano otras herramientas que permitirán complementar nuestro análisis del programa naturalista sobre las emociones.

Contra el lenguaje privado. Cuando nos preguntamos por nuestros conceptos mentales tendemos a pensar que ellos designan vivencias y eventos que solo nosotros, en una suerte de fuero interno, podemos percibir. Por ejemplo, cuando hablamos de nuestras sensaciones, creemos que ellas son algo extraño que ocurre solo de manera privada a la que únicamente nosotros tenemos acceso. Así, elucidar los conceptos de la psicología (como el concepto de 'sensación') requeriría una especial introspección que nos devele lo que es característico de nuestras vivencias. Esto puede llevarnos a pensar que desconocemos de manera objetiva nuestra vida mental y que, anclándolo a las observaciones anteriores, necesitamos de una ciencia que supere esta barrera. Esta imagen de la mente como un halo privado al que yo, en tanto único sujeto de mis experiencias, tengo acceso directo, proviene especialmente de la filosofía cartesiana, aunque James también acude a ella para sostener que la introspección ha de ser el principal método en psicología.[54]

Uno de los aportes más reconocidos de parte de Wittgenstein ataca justamente esta imagen de lo mental. Este famoso argumento se ha llegado a conocer como el "argumento contra el lenguaje privado". Este procede, a muy grandes rasgos, de la siguiente manera: Recordemos que

[54] En el caso de James, sin embargo, la introspección es falible. En el caso de Descartes, es —en principio— infalible.

Wittgenstein nos ha ofrecido una imagen del lenguaje que intenta separarse de la idea de una esencia subyacente al lenguaje, y mencionamos de manera muy breve que esta imagen que el autor critica supone que el lenguaje tiene una función exclusivamente referencial. En otras palabras, Wittgenstein quiere criticar la pretensión de que la única función del lenguaje es describir estados de cosas, lo que llevaría a que el significado se equipare a la referencia a objetos. Puesto de otra manera, bajo la imagen de que la función del lenguaje es únicamente referir, significado y objeto son una y la misma cosa.

En contraste con esta imagen, Wittgenstein nos invita a ver en el lenguaje una gran variedad de usos similares a un juego. En el lenguaje, como en un juego, hay reglas de uso, criterios de corrección y prácticas en la que se usan las palabras según estas reglas. Ellas no tienen por qué ser estrictas; las reglas no necesitan estar a manera de manual de instrucciones que leemos antes de usar el lenguaje. Ellas están íntimamente ligadas a nuestras "formas de vida". *Grosso modo*, esto significa que usamos y aprendemos el lenguaje en la práctica con los hablantes, en la forma en que vivimos en comunidades que usan las palabras y las oraciones de uno u otro modo.

Ahora bien, es importante notar que hemos dicho que las reglas del lenguaje tienen criterios de corrección. Si digo "La mesa ladró", parece que estoy usando mal la palabra 'mesa' o el predicado 'ladrar'; la mesa no es algo que ladre. En este sentido, decimos que no estoy siguiendo la(s) regla(s) de uso de esta palabra. ¿Cómo sabemos esto?

Porque hemos aprendido en una comunidad en que 'mesa' se usa en ciertos contextos, y este (en el que digo "La mesa ladra") no es uno de ellos, es decir, esta no es la manera en la que hablamos. De ahí la afirmación de que los criterios de corrección del lenguaje sean públicos, toda vez que requieren de una comunidad de hablantes.

Sin embargo, a la hora de aplicar esto a nuestros estados mentales surgen varios problemas. ¿Cómo sabemos, por ejemplo, que lo que nosotros llamamos 'dolor' es lo mismo que otros llaman por ese nombre? ¿Cómo sabemos que nos referimos a lo mismo con las palabras 'sensación', 'emoción', 'pensamiento', etc.? Dado que tendemos a pensar que estas palabras designan eventos de nuestra vida privada, estamos tentados a creer que nosotros sabemos identificar nuestras vivencias (privadas) con esos nombres, y el problema surge porque no tenemos acceso a las vivencias del otro. En otras palabras, creemos saber lo que es para nosotros "sentir dolor", por ejemplo, pero no qué es el dolor para los demás. En ese orden de ideas, pareciera que tuviéramos criterios de identificación privados para nuestras vivencias, de modo que sabemos identificar qué es el dolor gracias a un acceso privilegiado a nuestra mente. De ser así, tendríamos una suerte de lenguaje privado que nos permitiría identificar los conceptos mentales; precisamente, nuestra dificultad surge cuando intentamos aplicar estos conceptos a otras personas y notamos que no tenemos clara la forma de aplicación correcta, toda vez que no sabemos qué clase de vivencias tienen los demás.

Con esto en mente, preguntémonos qué sería un lenguaje privado. Este sería un lenguaje que un hablante posee sin que nadie más participe de él, toda vez que es un lenguaje que identifica únicamente sus propios estados mentales privados. En este sentido, este lenguaje tendría reglas cuyos criterios de corrección solo conocería quien lo posee; de lo contrario, no sería completamente privado. En principio, nosotros —como únicos hablantes de nuestro lenguaje privado— sabríamos cuáles son las reglas de su uso y cuáles los criterios para decir que seguimos de forma correcta estas reglas. Sin embargo, dice Wittgenstein:

> [...] 'seguir la regla' es una práctica. Y *creer* seguir la regla no es seguir la regla. Y por tanto, no se puede seguir 'privadamente' la regla, porque de lo contrario creer seguir la regla sería lo mismo que seguir la regla (IF §202, énfasis en el original).

Este pasaje y muchos otros similares quieren decir que para que exista tal cosa como "corrección" en el uso de una regla, debemos poder distinguir entre un hablante que usa correctamente la regla y otro que solo *cree* usarla correctamente. En un lenguaje privado no podríamos hacer esta distinción. En él no podemos determinar qué es seguir correctamente la regla, qué es creer que se sigue correctamente, o qué es equivocarse en el uso del lenguaje. En palabras del autor: "Se querría decir aquí: es correcto lo que en cualquier caso me parezca correcto. Y esto sólo quiere decir que aquí no puede hablarse de 'correcto'" (IF §258).

Para elaborar mejor el argumento, tomemos un ejemplo ampliamente discutido por Wittgenstein a lo largo de su obra: el caso de los colores. Supongamos por un momento que para identificar los colores me valgo de una tabla con distintas muestras de color, donde cada muestra tiene una etiqueta que dice “verde”, “rojo”, ”azul”, etc. Esta tabla solo es accesible a mí; nadie más puede ver mi tabla de muestras. Supongamos que alguien me pide que le pase el lápiz azul. A continuación, yo saco mi tabla de muestras, voy color por color hasta que llego a la etiqueta que dice “azul”, y tomo el lápiz cuyo color coincide con la muestra. Ahora, para dar un giro a nuestro caso, supongamos que desaparecen todos los objetos azules del mundo (salvo nuestra muestra). ¿Qué uso podría tener la muestra azul en mi tabla de muestras? Si todos los objetos azules desaparecen, ya no tengo ningún uso para esta muestra. No obstante, dado que todavía tengo la muestra, podríamos pensar que la muestra queda allí a la espera de poder ser utilizada. En otras palabras, la regla sobre cómo usar “azul” (buscando en la tabla de muestras aquella muestra que coincida) sigue a nuestra disposición, incluso cuando públicamente no tenga ningún uso. Ella se convierte ahora en una regla a la que solo yo puedo acceder, i.e., una regla privada.

Digamos ahora que alguien toma nuestra tabla sin que nosotros lo sepamos, y cambia la muestra azul por una muestra roja. Ahora, ¿cómo podemos saber nosotros que la tabla ha sido modificada? Si seguimos a Wittgenstein, no tenemos manera de saberlo. En ese sentido, llamaremos a los lápices rojos “azules”, siguiendo nuestra regla de buscar en nuestra

tabla de muestras. Creeremos, entonces, que seguimos bien la regla y, en el fondo, estaríamos siguiendo correctamente la regla, pues en efecto la etiqueta "azul" coincide con una mancha roja. De allí que Wittgenstein diga que creer seguir la regla es lo mismo que seguir la regla y que esto significa que no podemos hablar aquí de lo correcto. En un caso como este no podemos salir de nuestro error, pues no contamos con ningún criterio para hacerlo. Lo "correcto" sería ahora hablar de "azul" cuando veamos objetos rojos.

Este ejemplo debe es más bien mostrarnos que los criterios de corrección del uso de un término no pueden ser privados, pues en ese caso no habría forma de determinar su corrección; en otras palabras, no serían criterios de corrección en absoluto. Los criterios para salir de nuestro error serían públicos, es decir, necesitaríamos de otros que corrigieran el uso que hacemos de la palabra "azul". Si ése es el caso, entonces la forma en la que hablamos de lo azul no es gracias a una tabla a la que solo podemos acceder nosotros, es en virtud de nuestra vida con otros, de nuestras formas públicas de habla, que determinamos la manera de usar los conceptos.

Ahora, la figura de la tabla con muestras a las que solo podemos acceder nosotros es similar a si postuláramos una vida mental privada sobre la cual determinamos los usos de los conceptos. Wittgenstein recomienda el siguiente método para aclarar la gramática de los conceptos psicológicos:

> Para nuestros propósitos, podríamos reemplazar perfectamente bien cada proceso de imaginar por un proceso

> de mirar un objeto o de pintar, dibujar o modelar; y cada proceso de hablar con uno mismo por procesos de hablar en voz alta o de escribir (CA, p. 31).

Este método, si bien aquí se enuncia solo al servicio de la imaginación o el hablar con uno mismo, puede ampliarse a otras áreas de la filosofía de la psicología. De hecho, este es un recurso frecuente que usa Wittgenstein para tratar temas referentes a nuestra vida mental. Siguiendo esta recomendación, hemos reemplazado un proceso de pensar en un color por un proceso de revisar en una tabla con muestras de color. Con esto en mente, podemos decir: si para determinar qué objeto tomar al recibir la orden "Pásame el lápiz azul" recurrimos introspectivamente a una especie de muestra mental, podremos construir el caso para obtener la misma conclusión. Podemos decir que si olvidamos nuestra muestra azul en el fuero mental y la confundimos con una roja, entonces caeríamos en el mismo error de antes: no podríamos saber que estamos en un error, a menos que los demás nos lo informen. En ese sentido, pensar que los criterios de corrección de un concepto están en mi vida mental privada carece de sentido, pues no podría decir cuándo uso correctamente un concepto y cuándo no. Los criterios de corrección son, de nuevo, públicos, y no privados, incluso para aquellos conceptos que remiten a mi vida mental.

La noción de 'criterio de corrección' es muy importante para la investigación de Wittgenstein. A lo largo de su obra, la palabra 'criterio' aparece en varios momentos,

con distintas acepciones y usos.[55] Entre las investigaciones relacionadas con los criterios, Wittgenstein hace una distinción muy útil, en especial para temas psicológicos, entre un criterio y un síntoma. Cuando pensamos, por ejemplo, que un concepto como 'dolor' refiere a una vivencia privada a la que los demás no tienen acceso, tendemos a pensar en los gritos de dolor como un síntoma de algo que ocurre en el fuero interno de la persona. Sin embargo, como intentaremos ver, en muchos casos estas expresiones fungen como criterios de aplicación de los conceptos.

Síntomas y criterios. Cuando hablamos de los estados mentales de otras personas podemos preguntarnos cómo sabemos que ellas se encuentran en tal estado. Por ejemplo, si digo que el otro tiene dolor de muelas, puedo preguntarme cómo sé esto. Puedo, a continuación, dar una serie de *criterios* para justificar mi afirmación; por ejemplo, la persona se lleva la mano a su cachete y grita. La noción de criterio es esencial para la filosofía de la psicología de Wittgenstein. Según él, la gramática de varios conceptos psicológicos puede aclararse atendiendo a los criterios con los que adjudicamos estos estados.

Son varias las maneras en las que Wittgenstein trata la palabra 'criterio'. De manera preliminar podemos decir que un criterio es una regla que nos dice cómo aplicar un concepto. Si digo que la silla mide un metro, el criterio para determinarlo es, por ejemplo, haciendo una medición con

[55] Para una discusión sobre los distintos usos de la palabra 'criterio', véase Albritton (1959).

una vara, esto es, lo que nos dice si podemos aplicar el predicado "mide un metro" a la silla es la vara. En el caso del dolor, puedo decir que otro siente dolor de muelas cuando lleva su mano a su mejilla y grita, i.e., estos comportamientos sirven para decir si podemos aplicar el concepto de dolor en este caso.[56]

No hemos de confundir, sin embargo, un criterio de algo con sus "síntomas". Para Wittgenstein, un síntoma es "un fenómeno que la experiencia nos ha enseñado que coincide, de un modo o de otro, con el fenómeno que es nuestro criterio definidor" (CA, p. 55). Supongamos, verbigracia, que una roca cae sobre mi brazo, y a continuación siento un terrible dolor. Sabemos que este dolor puede ser un síntoma de una fractura; pero los criterios para identificarla son distintos. Por ejemplo, podemos hacer una radiografía y ver que hay un hueso roto. El dolor coincide con la fractura, pero no la define de ninguna manera.

De aquí, podemos decir que un criterio, al ofrecernos una regla de aplicación de un concepto, es "una forma descuidada de establecer una definición" (CA, p. 55). Al exponer esta distinción entre síntomas y criterios, Wittgenstein nos advierte que es común que lo que antes era un criterio pase

[56] De este tipo de afirmaciones varios comentaristas afirman que Wittgenstein despliega una forma de conductismo. Esto, sin embargo, no reconoce que Wittgenstein no desea hacer una teoría sobre la mente. Afirmar que Wittgenstein es conductista sería acusarlo de formular una teoría, una en la que los estados mentales refieren a conducta. Esto, por supuesto, contradice la forma misma en la que Wittgenstein quiere practicar filosofía, de modo que constituye una mala lectura del autor.

luego a ser solo un síntoma. Podemos pensar, por ejemplo, que un criterio definidor de una infección era la fiebre; en la actualidad la fiebre no es más que un síntoma de la infección y tenemos ahora otro tipo de criterios (e.g. exámenes médicos) para determinarla. En ese sentido, decir cosas como "siempre que hay una fractura hay un hueso roto" es una tautología, en tanto que está enunciando la definición misma —o mejor, el criterio definidor— de la fractura. En cambio, decir "siempre que hay una infección, hay fiebre" es hacer una hipótesis, ya que no se está enunciando ninguna definición. Allí se están relacionando dos fenómenos con definiciones distintas de los que es lógicamente posible decir que no tienen relación en todos los casos. En otras palabras, dar un síntoma es dar una hipótesis que podemos corroborar o abandonar; dar un criterio es, en algún sentido, una definición.[57]

Hemos dicho que Wittgenstein, al acercarse a la psicología, hace uso de esta distinción para no caer en la ilusión de que los conceptos psicológicos refieren a estados privados de una mente. ¿Cómo funciona aquí esta herramienta? Wittgenstein constantemente se pregunta por los "criterios" de aplicación de los conceptos psicológicos, de modo que

[57] Aquí puede haber cierto parentesco con la distinción entre enunciados analíticos y sintéticos. No obstante, es necesario advertir que Wittgenstein quiere criticar esta manera de hacer filosofía y este tipo de distinciones, pues esto supone hacer una teoría sobre los enunciados y el significado. Eso es, precisamente, lo que se quiere evitar. Siguiendo esta línea, la invitación al lector es a no introducir aquí distinciones de este tipo, incluso cuando puede haber alguna similitud.

los primeros nos ofrezcan reglas de aplicación públicas de los segundos. Si revisamos estos criterios, podremos ver que no necesitamos postular una vida privada interna para dar cuenta de cómo hablamos de nuestros estados mentales.

El ejemplo paradigmático de este tipo de problema es el ya mencionado caso del dolor. Hemos dicho que enunciados como "No puedo dudar de mi dolor" parecen sugerir un acceso especial a nuestra vida mental, y que ella —la vida mental— solo está disponible en nuestro fuero privado. Además, esto nos puede llevar a pensar que no podemos saber con certeza si el otro siente dolor, o qué sería el dolor para esa persona. Cuando hablamos del dolor de otros, podemos pensar que estamos haciendo solamente una conjetura, pues no tenemos ningún acceso epistémico a la mente del otro, como sí lo tendría a mi propia mente. Apliquemos la noción de criterio a este caso.

Si nos preguntamos cómo adjudicamos dolor a otros, vemos que nuestros criterios son similares a los contemplados por Wittgenstein y citados anteriormente: la persona grita y lleva su mano a su mejilla. Si bien es posible que la persona haga esto sin sentir dolor (e.g. en una obra de teatro o para engañarnos), nosotros aprendemos a usar este vocablo gracias a este tipo de conducta. Eso no significa —como piensan algunos— que la conducta *sea* el dolor; la conducta es solo nuestro criterio para dar con el dolor de otros. Por ejemplo, si la otra persona grita y lleva su mano a su mejilla y nos dice "¡Me duele mi muela!", no tiene sentido preguntarle "¿Estás seguro?". En contextos ordinarios, atribuimos a esta persona el dolor en virtud de

sus expresiones (su grito, su gesto con la mano y su declaración). Ellas fungen aquí como los criterios que nos dicen qué es que una persona siente dolor.

Podría objetarse aquí que, si bien puede que atribuyamos dolor en tales contextos, estos gestos no definen de ningún modo el dolor. Más aún, podríamos llevar a la persona a un consultorio médico y revisar si efectivamente la persona está adolorida. A esto podemos replicar: esta práctica de llevar a la persona al médico solo es posible porque hemos determinado criterios para la aplicación del concepto 'dolor'. En principio, podemos decir que el otro está adolorido sin llevarlo al consultorio; pero cuando lo llevamos al consultorio, estamos cambiando los criterios. En este caso tomamos como criterio un signo en una máquina y ya no las expresiones. De cualquier forma, lo importante es ver que esos signos, ahora funcionando como criterios, no nos ofrecen síntomas del dolor; ellos son, en cierto sentido, el "sentir dolor". En otras palabras, ellos son los que determinan qué es sentir dolor, i.e., los criterios de corrección del uso del concepto. Tomar los signos de la máquina como síntomas en este caso no tendría sentido e, incluso si lo hiciéramos, tendríamos que contar entonces con otros criterios para saber qué es lo que tenemos que buscar como causa de estos signos, de modo que no podríamos escapar a la asignación de criterios.

Notemos que cualquiera sea el conjunto de criterios que adoptemos, ellos no están sometidos a la confirmación empírica, como sí lo están los síntomas. Cuando asumimos un criterio para aplicar un concepto no tiene sentido que

hagamos experimentos o cosas por el estilo. Ellos determinan qué es lo que llamamos por tal y tal nombre. En cambio, sí puedo hacer experimentos sobre los síntomas de un fenómeno. En el caso anterior sobre la fiebre, puedo hacer experimentos para revisar si efectivamente siempre que hay una infección, hay fiebre. En este sentido, de nuevo, postular un síntoma de algo es hacer una hipótesis; postular un criterio es, en cierto sentido, formular una definición torpe del concepto. En ese sentido, que la conducta sea criterio del dolor no significa que el dolor refiera a la conducta; pero sí significa que nuestro criterio de aplicación de la palabra, i.e., nuestra manera de usar la palabra, es *público*, puesto que todos podemos ver la conducta de dolor. Así, podemos decir que no aprendemos a usar la palabra "dolor" porque tengamos una vivencia privada especial a la cual anclamos esta palabra. En cambio, aprendemos el uso de esta palabra porque hemos visto a otros presentar ciertas conductas como los gritos. Tras un adiestramiento en una comunidad lingüística hemos aprendido a usar la palabra 'dolor' cuando vemos este tipo de comportamiento (CA, p. 54).

Si seguimos esta línea de razonamiento, no tiene sentido pensar que las expresiones de dolor (entre ellas, el grito de dolor) son síntomas de un mecanismo oculto. Cuando pensamos así, intentamos destapar el misterio e investigar qué es lo que ocurre "realmente" cuando sentimos dolor. No obstante, si las observaciones de Wittgenstein surten efecto, debemos aprender a ver que no hay tal misterio, y que distinguir entre lo que parece ser un síntoma es, en muchos casos, un criterio de aplicación.

Este es solo un caso de aplicación de la noción de criterio como parece usarla Wittgenstein. Por supuesto, el análisis del autor es mucho más profundo que el presentado aquí. Sin embargo, el ejemplo sirve para notar cómo podemos ver en un concepto psicológico la poca necesidad de una vida privada a la cual referirlo. Podemos preguntar por los criterios y ver, a continuación, que los conceptos psicológicos no designan eventos privados, sino conductas. En otras palabras, lo que debemos aprender es que muchas de nuestras formas de habla sobre los conceptos mentales no son síntomas de algo privado que debemos develar. Algunas formas de habla —como los gritos de nuestro dolor— son criterios de uso de un concepto. El dolor solo nos aparece como misterio cuando pasamos por alto esta distinción entre criterios y síntomas y creemos que el dolor esconde una naturaleza oculta a ser develada.

De modo similar ocurre con las emociones. Buena parte de los misterios que surgen al pensar sobre nuestras emociones radica en esa tentación de pensar que las expresiones de emoción son síntomas de una vivencia interna especial. Si atendemos a estas observaciones, podemos quizás desprendernos de dicha tentación. Sobre este punto, y otros que hemos tocado brevemente aquí, nos detendremos en el siguiente capítulo.

4. Críticas a la naturalización de las emociones

4.1. Algunas confusiones gramaticales

Hasta este momento hemos hecho un despliegue de varias herramientas wittgensteinianas para el estudio de algunos conceptos psicológicos. De manera general hicimos un acercamiento, ya no científico a la luz de una metodología de reconstrucción racional, sino gramatical. Esta nueva perspectiva gramatical sobre los conceptos psicológicos tiene como finalidad disipar las confusiones que surgen cuando filosofamos sobre ellos. Entre estos conceptos encontramos el de las emociones. Pasemos ahora a aplicar estas herramientas al programa de naturalización de las emociones, mostrando algunas de las confusiones conceptuales que subyacen a este.

Retomemos nuestra definición del programa de naturalización de las emociones. Este programa, iniciado por Descartes, define las emociones como la percepción de reacciones corporales. Según este programa, la esencia de las emociones es, en el fondo, fisiológica. Este aspecto del programa es lo que define su núcleo firme. El avance de este programa tuvo como sus impulsores —entre otros— a James, quien introdujo la teoría de la evolución de Darwin para asignarle funciones a las emociones, y a Prinz, quien

insufló la teoría de la representación de Dretske para mostrar cómo puede explicarse la intencionalidad en términos naturalizados. Como vimos en el segundo capítulo, las emociones son intencionales en tanto hacen referencia a objetos por fuera de ellas. Incluso, ampliamos este análisis, siguiendo a Kenny, mostrando que podíamos hablar de objetos materiales y formales de la emoción, distinción que es central para la teoría de Prinz. A continuación, identificábamos también el problema que surge a la hora de tratar los objetos de nuestras emociones desde una perspectiva naturalizada, e identificamos algunos de los intentos con los que Descartes, James y Prinz atacaron el problema. En especial, señalamos que el programa giró en una dirección que permitió que Prinz, auxiliándose con la teoría de la representación mental de Dretske, mostrara cómo podemos hacer un análisis naturalizado de la intencionalidad de las emociones.

¿A qué responde la exigencia de un análisis naturalizado de la intencionalidad de las emociones? De manera resumida, el análisis naturalizado intenta ofrecer una descripción causal acomodada a las ciencias naturales de qué es lo que llamamos el objeto de nuestra emoción. En otras palabras, la naturalización de la intencionalidad nos pide dar respuesta a la pregunta "¿Por qué es X el objeto de mi emoción?". Frente a esta pregunta ofrecemos una respuesta siguiendo la teoría de Dretske: X es el objeto de mi emoción porque X causa ciertas reacciones corporales que luego llamo "emoción".

En el capítulo anterior notábamos cómo la metodología de Wittgenstein intentaba apartarse de este tipo de aproximaciones teóricas y nos invitaba a revisar el lenguaje ordinario,

usos corrientes de los conceptos en los que debemos ver distintas formas de vida. Con esto en mente, podemos comenzar una aproximación wittgensteiniana al problema de la naturalización de la intencionalidad estudiando esta exigencia por un análisis causal de la intencionalidad. Para ello estudiemos un caso posible en el que la intencionalidad de las emociones aparezca como misterio y sobre el cual estemos tentados a ofrecer una respuesta teórica. Imaginemos por un momento que un amigo llega a nuestra casa en lágrimas, e indagamos por su tristeza. Pensemos en el siguiente diálogo:

A: "¿Por qué esa cara? ¿Qué pasa?"
B: "Como puedes ver, estoy terriblemente triste".
A: "¿Y por qué tanta tristeza?"
B: "Es que mi perro ha muerto".
A: "¿Y por qué te sientes triste por la muerte de tu perro?"

En este diálogo, *A* interroga a *B* sobre su tristeza; en particular, *A* pregunta por qué *B* siente tristeza por la muerte de su perro. Según una teoría naturalizada de las emociones, debemos responder en términos de funciones y mecanismos que se disparan a causa de un estímulo sensorial. Diremos, por ejemplo, que el perro ha causado un estado mental, una representación que llamamos "tristeza".

Sin embargo, detengámonos un momento en la pregunta misma. Siguiendo la metodología de Wittgenstein, preguntemos cuál es el papel de ella. Con relación a las preguntas

"¿Por qué…?", vimos en el capítulo anterior la sugerencia de Wittgenstein de evaluar si estamos preguntando por las causas o si pedimos una razón. Supongamos ahora que nuestro amigo nos responde a continuación: "He tenido un estímulo sensorial particular después de ver a mi perro tendido en el piso sin respirar que ha desencadenado reacciones en mi cerebro que han agitado mi cuerpo de tal y cual manera". En este caso, nuestro interlocutor nos ha dado la *causa* de su tristeza. Sin embargo, esto no parece dejarnos satisfechos. Dicho así, no nos ha dicho nada sobre la relación que tiene la muerte del perro y su tristeza y sentimos que esto es algo que merece explicación. Como veíamos en el segundo capítulo, si nuestro amigo nos diera un listado con todas sus reacciones corporales, en principio esto no bastaría para una explicación —o identificación naturalizada— del objeto de su emoción. Nuevamente, la intencionalidad aparece como misterio, toda vez que la explicación causal —hasta ahora— no nos ha dicho nada sobre el objeto de su emoción, i.e., el hecho de que su perro ha muerto.

Dentro del programa de investigación que estamos examinando, es la teoría de Prinz la que nos ha ofrecido una explicación de la intencionalidad. Prinz, siguiendo a Dretske, nos ha propuesto ver relaciones causales y temas relacionales básicos en nuestras emociones. Así, nuestro amigo podría darnos una causa "completa" de su emoción, diciendo: "hay cierto objeto *O* que ha causado ciertas reacciones en mi cerebro, y dado que hay una relación contextual *C* entre *O* y mis sistemas representacionales, decimos que *O* es el objeto de mi emoción. En este caso, el sistema representacional que se

ha disparado es aquel que detecta una pérdida. Dado que ése es el tema relacional básico de la tristeza, y ha sido causado por el objeto *O*, que en este caso es la muerte de mi perro, entonces es por eso es que me siento triste." Nuevamente, esta respuesta suena algo extraña.

En este punto podemos ver que lo que perdemos de vista es el papel de la pregunta "¿Por qué…?" Si pensamos que debemos indagar por una causa, llegaremos a descripciones extrañas como la anterior, pues hemos confundido el rol que cumple la pregunta. En el diálogo que presentamos, *A* pregunta por una razón, no por una causa. En ese sentido, las respuestas de *B* no satisfacen la demanda de *A*. Si aprendemos a ver en esta pregunta la pregunta por una razón, no pediremos más explicación causal a propósito del objeto de nuestras emociones. Si de entrada nuestro amigo nos dice "Siento tristeza por la muerte de mi perro porque mi perro había estado conmigo desde niño", quedaríamos medianamente satisfechos con su respuesta. En principio, no buscaríamos mayor explicación.

Si insistimos en ver en la muerte del perro una causa de la emoción, podemos intentar aclarar el asunto de la siguiente manera. En las *Lecciones y conversaciones sobre estética, psicología y creencia religiosa,* Wittgenstein explora el caso de la expresión "Siento desagrado y conozco la causa" (refiriéndose a lo que alguien podría sentir frente a un objeto estéticamente desagradable).[58] Dice Wittgenstein:

---

[58] El caso particular que estudia Wittgenstein es el de una puerta que está mal hecha, y frente a la cual demandamos una corrección (como hacerla más

> Decir "conozco la causa" me trae a la mente el caso de la estadística o el rastro de un mecanismo. Si digo: "Conozco la causa" parece que hubiera analizado los sentimientos (como analizo el sentimiento de oír mi propia voz y al mismo tiempo frotarme las manos), cosa que, por supuesto, no he hecho. Lo que hemos dado es algo así como una explicación gramatical [al decir que el sentimiento está "dirigido"] (Wittgenstein, 1966/1992, p. II, §18; inciso en el original).

En el capítulo anterior vimos que ofrecer una causa es dar una hipótesis susceptible de confirmación empírica. Sobre las causas tiene sentido hacer experimentos, o siguiendo esta cita podemos hacer análisis estadísticos o rastrear el funcionamiento de un mecanismo. Pero cuando decimos que nuestro sentimiento está dirigido, no estamos dando el rastro de un mecanismo. No hacemos experimentos para verificar si este es el objeto de nuestra emoción o no. Imaginemos que cuando nuestro amigo nos dice que está triste por la muerte de su perro, respondemos "¿Estás seguro? Hagamos una serie de experimentos para confirmar eso que dices". En este caso parece que sucede algo extraño. Pareciera que no estamos jugando el juego del lenguaje que se está presentando aquí, pues no estamos dando la respuesta que se espera de nosotros en este contexto.

---

alta o más baja). Es sobre esta sensación de incomodidad que Wittgenstein estudia la expresión "Siento desagrado [por la puerta] y conozco la causa". Véase Wittgenstein (1966/1992, II, §§13-16).

Si atendemos a las observaciones de Wittgenstein, debemos decir que, cuando decimos que la tristeza está dirigida a la muerte de mi perro (que la muerte de mi perro es el objeto de mi tristeza), estamos haciendo una aclaración sobre las demandas de un juego de lenguaje, a saber, aquel en el que damos razones y aclaramos el objeto de nuestras emociones. Con esa respuesta no estamos dando causas sobre mi tristeza, de modo que un análisis causal parecería no tener lugar. En otras palabras, dar el objeto de mi emoción no es dar un hecho susceptible de confirmación; es, en cambio, una aclaración sobre el uso del lenguaje en un contexto.

Como vimos en el capítulo anterior, en algunos pasajes Wittgenstein compara dar razones con dar el proceso de un cálculo. Cuando ofrecemos el proceso mediante el cual hemos calculado algo, no estamos ofreciendo las causas de nuestro cálculo. Estamos, si acaso, dando una justificación[59] de nuestro resultado (1966/1992, III §15). Esta justificación, en cierto sentido, nos aclara cómo hemos seguido una regla, esto es, cómo hemos usado un determinado concepto en un juego con criterios delimitados. En ese sentido, al dar razones de nuestra tristeza, aclaramos el objeto de nuestra emoción y —en el fondo— aclaramos el uso del concepto de

[59] Aquí, justificación debe tomarse en un sentido amplio. En general, se entiende por "dar una justificación" el dar razones por las cuales se da una acción. Si bien podría discutirse a continuación si estas razones son *buenas* razones o no, de modo que podría discutirse sobre la dimensión normativa de una justificación, esto no sería objeto de nuestro trabajo. Por lo pronto, permítase anotar que dar una justificación es dar una razón, independientemente de los posibles criterios intersubjetivos de su validación. Si el lector lo desea puede omitir la palabra 'justificación' y pensar únicamente en el acto de dar razones.

tristeza en aquellos contextos en los que surge la pregunta (como el del dialogo anterior). Para decirlo de otra manera, con las razones que damos sobre nuestra tristeza estamos mostrando que el uso que hacemos del concepto de tristeza se adecua a las demandas de un juego de lenguaje particular. Con ello, nuevamente, no damos causas ni esperamos un análisis naturalizado; al contrario, quedamos satisfechos con una respuesta como "Mi perro estuvo conmigo desde niño".

Podríamos continuar la insistencia filosófica y objetar: ¿por qué dices que quedaríamos satisfechos con esa respuesta? ¿En qué se fundamenta esa idea? En este punto, podríamos responder que por lo general, esta es una de las formas en las que hablamos y respondemos sobre nuestras emociones. Este es un caso ordinario, cotidiano, que bien podría presentarse en una situación concreta. Sin embargo, el filósofo tradicional podría quedar insatisfecho, buscando una fundamentación más precisa de por qué damos respuestas de este tipo. No obstante, es aquí donde la investigación wittgensteiniana debe detenerlo, y debe mostrarnos así hablamos, esto es, que esa es la práctica que subyace a este tipo de preguntas como "¿Por qué te sientes triste por la muerte de tu perro?". En lugar de intentar ver una razón profunda para sostener esto, deberíamos recordar que esa es la manera en la que hablamos de nuestras emociones.

Si tenemos razón en nuestro análisis, podemos ver uno de los errores que surgen al interior del programa de naturalización de las emociones. Pensar que todo reporte sobre mis emociones debe reducirse a lenguaje fisiológico nos lleva a confundir el papel que cumplen preguntas como la

anterior. Tras esta confusión, pensamos que es un misterio cómo un objeto en el mundo puede *causar* una emoción en mí y a continuación desplegamos toda una teoría naturalizada de la intencionalidad para que nos auxilie. En este sentido, la discusión sobre la intencionalidad que hemos rastreado desde Descartes hasta Prinz se basa en una confusión de este tipo. Para este programa, la intencionalidad de las emociones aparece como misterio —en parte— porque la reducción a lo fisiológico nos lleva a pensar en las preguntas "¿Por qué…?" como preguntas causales, frente a las cuales los objetos de mi emoción son algo extraño que requiere explicación, como si estos objetos fueran un truco de un mecanismo que debemos develar.

Ahora bien, ¿por qué insistir en el lenguaje fisiológico y generar esta confusión? ¿Por qué no detener el programa en el problema de la intencionalidad, y decir que esta dificultad falsea de suyo la tesis central del programa? ¿Por qué insistir en la necesidad de una definición fisiológica sobre las emociones? Aquí podemos atrevernos a decir que el programa de investigación en cuestión ha impuesto una definición que, de tomarse como una definición general, lleva a estas confusiones, a saber, la definición de las emociones como la sensación de reacciones corporales.

Al imponer esta definición como el núcleo firme del programa todos los avances de la investigación deberán conservarla y reforzarla. Esta definición, podemos decir, fija un criterio de uso para el concepto de emoción. Al acercarnos a la noción de criterio, dijimos que —para Wittgenstein— un criterio es una forma torpe de establecer una definición.

Así, es legítimo hacer el paso y decir que esta definición fija criterios de uso.[60] En este caso, la investigación estará encaminada a buscar reacciones corporales para identificar con ellas a las emociones. Dado que esta pesquisa exige un lenguaje causal propio de las ciencias naturales, pensamos que tenemos que acomodar los objetos de las emociones a este lenguaje, y es allí donde la intencionalidad aparece como un obstáculo a superar. No obstante, podemos intentar desprendernos de esta confusión si aprendemos a ver una variedad más amplia de criterios para la atribución de emociones, más allá de las reacciones corporales. El problema está en que —como filósofos— nos aferramos a esta definición, pensando que Descartes, James y Prinz han hecho un descubrimiento filosófico importante, e intentamos generalizar esta definición y aplicarla inmediatamente a casos en los que no tiene lugar. Si aprendemos a ver en esta definición la fijación de criterios y no una hipótesis que hemos descubierto, quizás podamos detener esta generalización excesiva.

Tomemos un caso discutido por Wittgenstein y construyamos a partir de allí: "Hay *una* cosa de la que no puede decirse ni que es de 1 metro de longitud ni que no es de 1 metro de longitud, y es el metro patrón de París" (IF §50). El metro patrón de París, en cierto sentido, mide un metro, pero no porque tenga la propiedad física de medir un

---

[60] Incluso en el primer capítulo decíamos que la metodología de Lakatos nos permite incluir tesis filosóficas en el quehacer científico. En este caso, una definición que nace en el marco de la discusión filosófica dictamina los cursos de la investigación sobre las emociones.

metro, sino porque esta vara nos sirve de herramienta para modelar el uso de "un metro" (y toda la métrica). En otras palabras, la vara que se encuentra en París no tiene una propiedad oculta a la que llamemos "un metro", sino que con ella hemos convenido lo que significaría la noción de "un metro". Así, decir "El metro patrón de París mide un metro" no es más que ofrecer el criterio de uso del concepto 'metro'. En este sentido, esta proposición ("El metro patrón de París mide un metro") no nos da una hipótesis a confirmar. Supongamos que un grupo de físicos estudia con cuidado esta vara, la miden con otras cintas métricas, hacen un listado de sus propiedades físicas y concluyen que esta vara mide un metro. ¿Han hecho un hallazgo novedoso? Dado que la vara modela el uso de "un metro", no vemos la necesidad de hacer análisis científicos a propósito de esta vara para descubrir esta propiedad oculta en ella. Los científicos, incluso al hacer todas estas mediciones, ya están inmersos en un juego de lenguaje en el que la noción de "metro" ha sido modelada con esta vara. Nuevamente, la definición ofrece un criterio de uso y no una contingencia empírica que debemos descubrir, esto es, un síntoma. Para decirlo de otra manera, el hecho de que cuando hagamos experimentos la vara en París mida un metro no es un síntoma de una propiedad especial, pues ella es de suyo el criterio para lo que llamamos "medir un metro".

Tomemos ahora una proposición similar: "☺ expresa felicidad". Tenemos, primero, un símbolo que designa una expresión y a continuación una afirmación sobre lo que ella expresa. ¿Qué estamos diciendo con esta proposición?

Podríamos pensar que estamos señalando una relación accidental entre "☺" y "felicidad"; a continuación, haríamos toda una cantidad de estudios para verificar si efectivamente la expresión ☺ designa felicidad. A propósito del programa de investigación del que nos hemos ocupado, es precisamente esta la vía por la que se investiga. Para este programa, las expresiones de las emociones son efectos de algo oculto que ocurre en el cuerpo y a continuación debemos investigar qué es eso que ocurre allí.[61] Más aún, para dicho programa esto que está oculto es algo que sentimos de manera privada, es una reacción corporal que solo se siente si estamos registrando de alguna manera lo que ocurre en nuestro cuerpo. Esto fortalece la idea de que las emociones son misteriosas y que merecen un estudio naturalizado para entenderlas.

En este punto, podemos notar que el carácter privado de mis emociones aparece en la teoría bajo la forma de un acceso especial que tengo a mis sistemas representacionales. Para decirlo de otra manera, el programa naturalista nos intenta convencer de que nuestras emociones son estados representacionales y dado que solo nosotros podemos acceder a ellos, las emociones son estados privados misteriosos. Dado este carácter privado de mis emociones, este programa nos lleva a pensar que las expresiones de emoción son síntomas de algo más que ocurre, a saber, de un estado representacional al que solo yo tengo acceso.

---

[61] Muy especialmente, este tipo de estudio fue inaugurado por Darwin en su obra *The Expression of Emotions in man and animals.* Véase Darwin (1998).

Al tratar el argumento contra el lenguaje privado, Wittgenstein nos ha intentado mostrar que no necesitamos la postulación de una vida privada interna para entender nuestros conceptos psicológicos. Si las reglas de aplicación de los conceptos son públicas, entonces debemos dirigir la mirada a las formas públicas de uso del concepto que nos ocupa. En otras palabras, en lugar de apegarnos a una vida mental interna a la que refiere el concepto 'emoción', podemos intentar ver cómo usamos el concepto en nuestra vida pública. Así las cosas, podemos formular la pregunta, ¿cómo identificamos emociones en otros?

En muchos casos, identificamos emociones en otras personas gracias a sus expresiones faciales. Cuando vemos a una persona sonreír, decimos "Esa persona siente felicidad". En el caso anterior, cuando nuestro amigo llega en lágrimas a nuestra casa, decimos "Esa persona siente tristeza". En estos casos, una expresión como "☺ expresa felicidad" puede pensarse como aclarando un criterio de uso del concepto, del mismo modo en el que "El metro patrón de París mide 1 metro" lo hace. Podemos hacer una gran cantidad de experimentos para saber qué ocurre siempre que tenemos la expresión "☺"; pero ellos no nos dirán nada nuevo sobre el concepto de felicidad. Pensemos en un niño que está aprendiendo a hablar de sus emociones, y le decimos que "☺ expresa felicidad". Para el niño, esa proposición modela el uso de un lenguaje, mas no es una hipótesis empírica a confirmar. El niño aprenderá a usar la palabra "felicidad" cada vez que vea la expresión "☺" en una persona, lo que lo introducirá en nuestra comunidad de hablantes. Luego,

el niño aprenderá a sustituir la expresión "☺" por otro tipo de expresiones, como "¡Qué alegría!" y otras similares. En este caso, la proposición en cuestión no designa una hipótesis empírica; la expresión ☺ no es un síntoma de felicidad, sino un criterio del uso de un concepto. En este contexto, ningún experimento podría darnos luces sobre cómo usar el concepto de felicidad. Para decirlo de otra manera: podemos darle sentido a la proposición "☺ expresa felicidad", incluso si no hemos hecho ningún análisis naturalizado de las emociones, toda vez que podemos usar esta proposición como la formulación de una regla. Así, la proposición no nos estaría diciendo algo sobre el mundo, sino simplemente sobre el uso de la palabra "felicidad" y del símbolo "☺".

Un análisis similar también puede correr con otras expresiones de emoción. Entre otras, la proposición "Siento tristeza por la muerte de mi perro" puede funcionar como la expresión de una razón sobre mi tristeza, una respuesta ajustada a una exigencia de la gramática de nuestro lenguaje, y no como síntoma de algo más que ocurre detrás. Esta proposición, en contextos cotidianos, ya nos sirve de criterio para la atribución de tristeza. El problema surge cuando pensamos en ella como refiriendo a un síntoma y a continuación intentamos indagar qué es lo que realmente ocurre detrás de la expresión. En otras palabras, creemos que la expresión es un síntoma de un sentimiento privado interno de la persona. ¿Por qué llegamos a pensarla como síntoma de algo más? Volviendo a lo dicho antes, podemos decir que esta confusión se produce porque en el interior del programa de naturalización de las emociones pensamos

que los únicos criterios para hablar de las emociones son los criterios fisiológicos. Así las cosas, lo que debemos ver es una variedad más amplia de criterios, de modo que los fisiológicos sean —a lo sumo— solo una parte de ellos.

En este punto podría objetarse: cuando atribuimos emociones a otros por medios que no son los estados corporales (como su expresión facial o la declaración "Siento tristeza por la muerte de mi perro") estamos conjeturando que la persona tiene tales y cuales estados corporales, aún si no los tenemos presentes (e.g., mediante alguna máquina que revise nuestros estados fisiológicos); cuando le atribuyo tristeza a una persona al ver su expresión le estoy atribuyendo un estado corporal. A ello podemos responder varias cosas. Primero, cuando le atribuimos a la persona un estado corporal, estamos haciendo una hipótesis: el enunciado "si tiene tristeza, tiene tal estado corporal" es una conjetura. Sin embargo, esta solo tiene sentido cuando yo acepto que la persona siente tristeza y a continuación conjeturo que debe tener tal o cual estado corporal. Incluso, tiene sentido, a continuación, conectarla a un monitor y verificar mi hipótesis, esto es, conducir un experimento.

No obstante, cuando yo le atribuyo tristeza (o temor en el caso de la siguiente cita) en un primer momento —sugiere Wittgenstein— no estoy haciendo una conjetura:

> En general, yo no conjeturo el temor [u otra emoción] en él: lo *veo*. No siento que yo esté deduciendo la probable existencia de algo interior {en él} a partir de algo exterior; es, más bien, como si el rostro humano fuera

> algo casi translúcido y como si yo lo viera no en una luz reflejada, sino en la luz que le es propia (OFP II, §170).

Siguiendo a Wittgenstein, mi atribución de temor es —en cierto sentido— directa. No conjeturo que él tiene temor, lo *veo.* Y aquí "Lo veo" significa: la expresión de temor no es un síntoma de algo más, sino un criterio de uso del concepto. Para tomar nuevamente el caso de la muerte del perro, la expresión, sea el llanto o la declaración "Siento tristeza por la muerte de mi perro", es en ciertos contextos el criterio con el que atribuimos tristeza a otra persona, mas no es la construcción de una hipótesis sobre su estado fisiológico. Las hipótesis solo tienen sentido una vez estamos inmersos en una práctica, en este caso la de atribuir emociones.

Por otra parte, aprendemos qué cuenta como expresión de una emoción porque estas muestras (el llanto o la declaración) nos enseñan cómo hablar de nuestras emociones, mas no porque hayamos hecho un análisis naturalizado y hayamos aprendido que las expresiones a veces acompañan ciertas reacciones corporales. Pensemos nuevamente en el caso del metro patrón de París. Cuando decimos que sabemos que el metro patrón de París mide un metro, no estamos diciendo que hayamos hecho muchos experimentos que nos han revelado la naturaleza de una vara determinada. Lo que estamos diciendo cuando declaramos saber que esta vara mide un metro es que reconocemos que la vara modela el significado de "medir un metro", esto es, reconocemos el papel que cumple la vara en nuestra gramática. Lo mismo ocurre aquí con las expresiones de emoción. Ellas

modelan nuestros usos de los conceptos de emoción, de manera análoga a como la vara modela nuestra métrica. Así, las expresiones se convierten —en muchos casos— en los criterios de atribución de emociones, en lugar de un efecto o síntoma de un mecanismo oculto, de una representación mental que ocurre detrás de esas expresiones.

Si insistimos en nuestra investigación filosófica, e indagamos por qué modelamos nuestros conceptos de emoción a partir de prácticas como las de identificar ciertas expresiones, veremos que no hay una respuesta teórica para sustentar esta declaración. ¿Por qué identificamos a la tristeza con una expresión como "☹"? ¿Qué determina que "☹" sea la expresión de tristeza y no de felicidad? Frente a estas preguntas, creemos que la evidencia biológica nos puede auxiliar, punto en el cual hacemos experimentos y observaciones como las de Darwin a propósito de las expresiones de las emociones. Sin duda esto puede arrojar datos interesantes, pero dicha investigación no nos dice qué es eso que llamamos "tristeza".

Podríamos seguir preguntando por el fundamento de nuestras expresiones emocionales, pero al final nuestra única respuesta sería: porque así hablamos, porque así convenimos que la expresión ☹ designa tristeza, o ☺ expresa felicidad, etc. En el fondo, el hecho de que asociemos ciertas formas de expresión y cierto tipo de preguntas con nuestras emociones responde a una práctica primitiva, lo que podríamos llamar un *protofenómeno*. No se trata aquí de que las expresiones sean síntomas de algo más, sino de que nuestra asociación entre ciertas expresiones y ciertos conceptos sobre nuestras

emociones es una manera de organizar nuestro lenguaje, una forma de práctica en nuestra comunidad lingüística. Por ello, expresiones como "☺ expresa felicidad" no indican, como hemos advertido, hechos del mundo que yo pueda confirmar mediante un experimento, sino una regla sobre cómo usar el concepto de felicidad. En el fondo, no hay una razón profunda que determine nuestra asociación entre expresiones y emociones; aquí no hay más que nuestra forma de vida.

Al aferrarnos a la definición o criterio propuesto por el programa de investigación sobre la naturalización de las emociones, confundimos lo que en algunos momentos es un criterio con un síntoma e intentamos hacer una investigación a propósito de estos supuestos síntomas. Esto, como ya notamos, produce la ilusión de necesitar una teoría que explique cómo eso que ocurre de manera privada —la sensación de reacciones corporales— toca o se relaciona de una manera extraña con un concepto. Volviendo sobre el argumento del lenguaje privado, la sensación se convierte aquí en una suerte de muestra privada que alcanza de manera misteriosa un objeto y esa es la ilusión de la que debemos desprendernos.

Si atendemos a estas recomendaciones, no nos sentiremos atormentados por una declaración como "Siento tristeza por la muerte de mi perro", pues ella no funciona como la declaración de un hecho privado, sino como una forma de aclarar mi lenguaje. Recordemos cómo surge el problema filosófico alrededor de la intencionalidad. Queremos hacer una descripción fisiológica de las emociones, pero en

principio ella no basta para explicar la intencionalidad. Sin embargo, como ya sugerimos, nuestras ansias por una descripción fisiológica nos llevan a pensar que las expresiones de emoción, entre ellas "Siento tristeza por la muerte de mi perro" son síntoma de eso que ocurre en mi cuerpo. Si aprendemos a ver que estas expresiones no fijan síntomas, sino que aclaran los criterios de uso, podemos detener la formulación del problema, pues ya no sentiremos ansiedad frente a esta declaración. Aquí debemos ver que cuando alguien nos informa que la muerte de su perro es lo que lo aqueja y produce su tristeza, no nos está dando un reporte sobre su estado mental interno y en esa misma vía no tiene sentido que ofrezcamos hipótesis sobre las relaciones causales entre el objeto de su emoción y su supuesto estado mental privado. Aquí una expresión como "Siento tristeza por la muerte de mi perro" sirve para aclarar las demandas de un juego de lenguaje en el que damos razones para nuestras declaraciones de emoción.

Cabe anotar que aferrarse a este criterio para identificar las emociones (las reacciones corporales) no es una cuestión injustificada o irracional. Hay varias razones para apegarnos así al criterio de las emociones como reacciones corporales. Por un lado, hemos visto que si atendemos a nuestra metodología de reconstrucción racional, es racional que los científicos de un programa se aferren al núcleo firme y muevan el cinturón protector para blindarlo. En este sentido, es de esperar —siguiendo nuestra reconstrucción— que Descartes, James y Prinz se apeguen a esta definición. Más aún, ella resulta interesante como proyecto científico,

aunque quizás no como teoría filosófica. Sobre esto profundizaré más adelante.

Por otro lado, en el segundo capítulo mencionamos que algunos sostienen que si es imposible naturalizar un término, entonces nuestro lenguaje sobre él carece de significado. En una defensa de la naturalización de la intencionalidad Moya (1994) sostiene, de la mano con Quine, que la imposibilidad de este proyecto puede llevar al eliminativismo:

> Lo que Quine nos ha hecho ver con claridad es que la autonomía e irreductibilidad de las propiedades intencionales es un arma de doble filo. Puede ser usada tanto para defender la importancia y sustantividad de las propiedades mentales que en la vida ordinaria atribuimos a los seres humanos como para apoyar la futilidad e inexistencia última de dichas propiedades y del supuesto conocimiento interpretativo que basamos en ellas (p. 228).

En ese sentido, dice Moya, naturalizar la intencionalidad puede ser una manera de salvar nuestro lenguaje mental de ser una simple ficción. Sin embargo, este es precisamente el error que debemos sacar a la luz: del hecho de que no podamos dar una definición en términos físicos no se sigue que se hable de una simple ficción. Para el caso de las emociones, este criterio se asume como una serie de reacciones corporales a las que debemos reducir las emociones, de lo contrario, estas parecerían ficciones. Pero recordemos que hay varios criterios, varias formas de hablar de las emociones y en este sentido no debemos pensar que la única forma

de determinar si podemos hablar legítimamente de algo es acudiendo al mundo de la ciencia natural.

Lo que necesitamos entonces no es una teoría científica que nos enseñe qué son las emociones; lo que necesitamos es una descripción de cómo hablamos de ellas, una descripción sinóptica que nos permita ver esta amalgama de usos y formas de vida que surgen en nuestra experiencia emocional. Podemos ver esto con una expresión de Wittgenstein en su primera obra:

> Sentimos que aun cuando todas las *posibles* cuestiones científicas hayan recibido respuesta, nuestros problemas vitales todavía no se han rozado en lo más mínimo. Por supuesto que entonces ya no queda pregunta alguna; y esto es precisamente la respuesta (Wittgenstein, 1921/1999, §6.52; énfasis en el original).

Podríamos hacer mil preguntas científicas sobre nuestra vida emocional, e incluso podríamos resolverlas todas. El programa de naturalización de las emociones podría enseñarnos todo lo que puede sobre nuestras emociones, y, sin embargo, hay hechos que quedarían por fuera, a saber, cómo usamos estos conceptos. Es aquí donde surge la necesidad de la descripción. La ciencia natural puede ir hasta cierto punto, pero hay asuntos que solo una descripción de nuestras formas de vida puede lograr. Aquí hemos realizado algunas aproximaciones en esta dirección y apenas nos hemos introducido en una jungla de enredos e ilusiones filosóficas. Hemos intentado ver, describiendo algunos

casos cotidianos, que no es claro que las emociones siempre se identifiquen con las reacciones corporales, que nuestros criterios de uso son más amplios y que la manera en la que hablamos de los objetos de nuestra emoción no deben parecernos un misterio científico a resolver. Por supuesto, es posible que haya muchas otras confusiones por tratar, pero si hemos tenido algo de éxito, podremos ver una primera manera de acercarnos a la naturalización de las emociones desde una perspectiva wittgensteiniana.

4.2. Anotaciones finales

En este trabajo intentamos usar algunas de las herramientas que ofrece Wittgenstein para sacar a la luz ciertas confusiones conceptuales que subyacen al programa de naturalización de las emociones. Para ello, definimos en el primer capítulo el programa de naturalización siguiendo la matriz historiográfica de Lakatos sobre los programas de investigación científica. Al aplicar esta matriz, veíamos como núcleo central del programa la definición de las emociones como la sensación de reacciones corporales y como movimientos del cinturón protector la introducción de la teoría de la evolución de Darwin (a propósito de introducir algunas explicaciones teleológicas) y la teoría de la representación mental de Dretske. Todo esto nos permitía elaborar una categoría general que reuniera las distintas teorías sobre las emociones alrededor de la empresa científica,y viendo como una de sus preocupaciones la naturalización de la intencionalidad.

Al abordar este tema en el segundo capítulo, señalábamos cómo las emociones son intencionales y qué tipos de problemas acarrea la intencionalidad para un programa naturalista. Como ejemplo tomamos el problema de la representación acudiendo al experimento mental de Putnam. Con esto en mente, volvimos sobre el programa de naturalización de las emociones para mostrar cómo el programa llegó a tener una teoría de la intencionalidad insuflada por Prinz a partir de la teoría de Dretske. En este punto, vimos que Descartes ya contaba con una vaga noción de representación y que James introdujo la noción de función, ambas protagonistas en la teoría de Dretske. Así las cosas, el programa de naturalización de las emociones, en su conjunto, tuvo el problema de la intencionalidad a lo largo de su desarrollo.

Luego, en el tercer capítulo, desplegamos algunas anotaciones de Wittgenstein sobre los conceptos psicológicos. Con ello, tomamos una perspectiva gramatical sobre el problema. En lugar de tomar el problema como un problema científico, dimos un paso atrás e hicimos algunas anotaciones sobre cómo pensar los conceptos psicológicos para evitar ciertas confusiones. Con esto, aplicamos algunas de estas herramientas al caso del programa de naturalización de las emociones, y vimos que las confusiones se desprenden de la definición que arroja el núcleo firme del programa.

Ahora bien, es hora de evaluar cuáles serían las consecuencias de nuestro trabajo. Si tenemos razón, ¿carecen de sentido los intentos por naturalizar las emociones? En otras palabras, ¿puede pensarse en una ciencia natural de

las emociones? ¿O debemos ceder ante las críticas y abandonar esta empresa?

La respuesta que ofrezco a continuación —de carácter tan solo tentativo— procede en dos partes. La primera consiste en que es importante considerar las dificultades conceptuales que acarrea tal empresa. Como ya dijimos, hemos señalado apenas algunas de las confusiones conceptuales subyacentes a este programa, pero quizás podamos encontrar muchas más. De ser así, pareciera que el programa de naturalización de las emociones hace malos usos del lenguaje en varias formas, entre ellas las tratadas aquí, de modo que, si seguimos a fondo las ideas que ofrece Wittgenstein, podríamos llegar a pensar que el proyecto de naturalización carece de sentido.

No obstante, la terapia wittgensteiniana no tiene como fin postular una teoría alterna que refute a su contrincante. En ese orden de ideas, no se trata aquí de decir que las teorías naturalistas de las emociones están absolutamente erradas, y que podemos ver este error a partir de una nueva teoría. Lo que debemos ver, como repetimos constantemente en el capítulo anterior, es la variedad de formas con las que usamos el lenguaje. Según esto, no debemos esperar una esencia general de las emociones, sino ver sinópticamente la amalgama de usos que están a nuestra disposición.

¿A qué va todo esto? Considero —y he aquí la segunda parte de mi respuesta— que sí podemos hacer ciencia natural sobre las emociones. Sin embargo, el error consiste en pensar que dicha ciencia está develándonos la naturaleza misma de este concepto. No debemos ver que la ciencia sea

una empresa sin sentido, sino que es una forma particular de uso del lenguaje que coexiste con muchas otras. Si hemos de decir algo sobre el papel de la ciencia en nuestras definiciones, será que la ciencia introduce usos técnicos particulares de términos que, en el caso de la psicología, usamos a diario. Pero estos usos técnicos no deben tomarse como más profundos que los demás. La ciencia es una práctica particular, con sus reglas y formas de proceder que no tienen por qué ser la última palabra sobre la realidad.

En ese orden de ideas, el programa de naturalización de las emociones está introduciendo o simplemente adoptando un uso diferente de los conceptos de emoción; no por ello está revelando su realidad oculta. Si este es el caso, no necesitamos de una teoría que concilie la ciencia con la vida cotidiana; estas son solo distintas áreas de nuestro lenguaje. Cuando los científicos hablan de las emociones quizás hablan de cosas distintas a cuando decimos sentir felicidad por llevar a cabo un proyecto importante. De ahí que el programa de naturalización, como ha sido desarrollado, incurre en el error de pensar que se trata de un solo lenguaje que la ciencia ha de depurar y que el filósofo debe generar teorías que ayuden al científico a lograr este fin. Tanto Descartes, como James y Prinz piensan que sus teorías ayudan a que el científico nos diga qué son las emociones; pero, como hemos dicho, aquí se pierde de vista el amplio espectro de formas de uso y se intenta dar una esencia última de las cosas.

Esta respuesta puede sonar confusa. ¿Cómo podemos concluir que es posible hacer ciencia natural sobre las

emociones, luego de todo un trabajo criticando el proyecto de naturalización? Aquí debemos recordar nuevamente el objetivo de la terapia wittgensteiniana: liberarnos de la confusión, no elaborar teorías. Si seguimos la metodología que nos hemos propuesto, debemos sacar a la luz los enredos del lenguaje. En nuestro caso, hemos atendido a las confusiones entre las razones y las causas y los síntomas y los criterios. Decir que por ello no podemos hacer ciencia de las emociones implicaría generar una teoría sobre cómo debe proceder la ciencia y esto yace por fuera de los objetivos de este trabajo. Aquí podemos recordar un famoso pasaje de las *Investigaciones* de Wittgenstein:

> La filosofía no puede en modo alguno interferir con el uso efectivo del lenguaje; puede a la postre solamente describirlo.
> Pues no puede tampoco fundamentarlo.
> *Deja todo como está* (IF, §124; énfasis fuera de texto).

Este parágrafo sugiere realmente que la filosofía, como la concibe Wittgenstein, no puede generar normas sobre cómo debe operar uno u otro juego de lenguaje. Es solo a partir de la descripción que se sacan a la luz las confusiones, pero estas descripciones no pueden a continuación normativizar sobre otros juegos, como el lenguaje científico. ¿Qué significa "dejar todo como está"? Significa no pretender que la filosofía modifique nuestros usos. En ese sentido, pretender que el trabajo filosófico que hemos realizado

modifique el quehacer científico sería contradictorio con la misma metodología que hemos adoptado.

Con ello podemos aclarar finalmente el resultado de esta investigación: no se trata de decir cómo hablar y cómo no hablar de las emociones, sino de reconocer que ciertas formas de habla escapan al proyecto unificador del programa de naturalización de las emociones. Es por ello que debemos ser cautelosos y no hemos de pretender que una teoría naturalizada capture todas las formas de uso del concepto de *emoción*. Es esa la conclusión a la que debemos llegar y no un rechazo total al trabajo científico que pueda haber sobre las emociones. La ciencia quizás tenga cosas que decir sobre nuestra vida emocional; lo que no puede hacer es decirnos cómo usar el concepto, esto es, no puede decirnos *qué* son las emociones.

En este sentido, todo el programa de investigación que hemos tratado en este trabajo es interesante justamente como empresa científica, mas no como empresa filosófica. El uso de la metodología de Lakatos nos permite rescatar algunos de los aspectos del programa, es decir, nos muestra que, como empresa científica, el programa puede dar frutos. Debemos, eso sí, tener presente que con él no buscamos una teoría general sobre qué son las emociones realmente, sino la formulación de un lenguaje que nos permita hacer ciencia sobre aquello que, antes de hacer ciencia, ya conocíamos bien. En otras palabras, quizás la idea de Descartes, James y Prinz de estudiar las emociones como la percepción de reacciones corporales sea provechosa para la empresa

científica, mas no para la empresa filosófica en la que pretenden funcionar.

¿Minan estas críticas todo intento de naturalización en general? Por lo pronto este es un camino de investigación abierto. Pareciera que la naturalización de los conceptos psicológicos sí supone algunas de las confusiones que hemos tratado aquí. A pesar de lo tentadora que pueda sonar esta tesis, es importante esperar a dar un juicio sobre el tema. Así como hemos criticado el proyecto de naturalización de las emociones por suponer un solo uso profundo del concepto, debemos abstenernos de dar generalizaciones apresuradas que sugieran una manera unívoca de usar la metodología wittgensteiniana. Si hemos dicho que hay variedad de usos de los conceptos de la psicología, también hay variedad de problemas y de proyectos, de modo que no podemos dar una tesis general a partir de observaciones locales como las presentes.

¿Podemos decir algo más sobre las emociones a partir de nuestras críticas? Ciertamente. Tener presente la variedad de usos nos permite recordar que no debemos pensar que la ciencia tiene la última palabra. En esta vía, podemos extender el análisis wittgensteiniano a posturas como la de Griffiths (1997) y sostener con él que no hay por qué pensar que las emociones forman una clase natural, mas no por las razones empíricas de este autor, sino porque hablamos sobre ellas de diversas maneras, no solo en el ámbito científico.[62] A su vez, podríamos intentar girar ahora hacia las

[62] Debo agradecer al profesor Miguel Ángel Pérez por darme la pista sobre esta idea.

teorías cognitivistas y mostrar que no en todos los casos es claro que las emociones puedan identificarse con juicios, como piensan Solomon (1973/1980) y Nussbaum (2001). Sin embargo, estos asuntos son materia de otro trabajo.

Finalmente, creo que con el ejercicio adelantado se abre un camino de investigación interesante. Muchos han tomado la intervención wittgensteiniana como una intervención negativa, esto es, siempre con el ánimo de destruir. Si bien aquí queremos disipar una ilusión, no por ello la intervención nos deja paralizados. Creo que podemos extender la intervención de Wittgenstein a otras áreas de la filosofía de las emociones, de la mente y de la ciencia cognitiva y con ello quizás tomar una perspectiva diferente sobre problemas como el atacado en este trabajo. En otras palabras, quizás sea interesante formular una perspectiva wittgensteiniana que, en lugar de destruir cualquier intento por elaborar un programa científico, lo aclare y evite el tipo de confusiones que hemos estudiado aquí.

El trabajo presente, finalmente, solo tiene la pretensión de dar luces sobre un proyecto, y sobre sus posibles problemas conceptuales. En el fondo, lo expuesto aquí debe dejarnos lo expuesto aquí es que las teorías de la emoción solo introducen o subrayan usos entre una variedad; pero no por ello nos dan una última definición.

Referencias

Albritton, R. (1959). On Wittgenstein's use of the term "criterion". *Journal of Philosophy, 56*(22), 845-857.

Aristóteles (n.d./2005). *Retórica* (Q. Racionero, Trad.). Madrid: Gredos.

Ayala, F. J. (2007). Darwin's greatest discovery: design without designer. En J. C. Avise & F. J. Ayala (Eds.), *In the light of evolution. Volume I: Adaptation and complex design* (pp. 3-22). Washington: The National Academies Press.

Baker, G. P., & Hacker, P. M. S. (1982). The grammar of psychology: Wittgenstein's bemerkungen über die philosophie der psychologie. *Language and Communication, 2*(3), 227-244.

Brentano, F. (1874/2009). *Psychology from an empirical standpoint* (A. C. Rancurello, D. B. Terrell, & L. L. McAlister, Trads. O. Kraus & L. L. McAlister (Eds.). Nueva York: Routledge.

Churchland, P. (1981). Eliminative materialism and the propositional attitudes. *The Journal of Philosophy*, *78*, 67–90. doi:10.2307/2025900

Cooper, W. E. (1990). William James' Theory of Mind. *Journal of the History of Philosophy, 28*(4), 571-593.

Damasio, A. R. (1994). *Descartes' Error*. Nueva York: Avon Books.

Darwin, C. (1998). *The expression of the emotions in man and animals*. Nueva York: Oxford University Press.

Davidson, D. (1963/2002). Actions, reasons, and causes. *Essays on actions and events* (pp. 3-20). Nueva York: Oxford University Press.

Davidson, D. (1970/2002). Mental events. *Essays on actions and events.* (2ª ed., pp. 207-224). Nueva York: Oxford University Press.

Descartes, R. (1633/1989). "Tratado de la luz". En *El mundo* (S. Turró, Trad.). Barcelona: Anthropos.

Descartes, R. (1628/1996). *Reglas para la dirección del espíritu* (J. M. Navarro Cordón, Trad.). Madrid: Alianza.

Descartes, R. (1649/2010). *Las pasiones del alma* (J. A. Martínez Martínez & P. Andrade Boué, Trads.) (2ª ed.). Madrid: Tecnos.

Descartes, R. (1887/1913). *Oeuvres de Descartes* (C. Adam & P. Tannery Eds.). París: Léopold Cerf.

Descartes, R. (1998/2004). *Descartes: The world and other writings* (S. Gaukroger Ed.). Nueva York: Cambridge University Press.

Descartes, R. (2007). *The correspondence between Princess Elisabeth of Bohemia and Rene Descartes (The other voice in early modern Europe)* (L. Shapiro, Trad., L. Shapiro Ed.). Chicago: University of Chicago Press.

Dretske, F. I. (1995). *Naturalizing the mind*. Cambridge: MIT Press.

Flanagan, O. (1991). Naturalizing the mind: The philosophical psychology of William James. *The science of the mind*

(2ª ed. rev. y ampliada, pp. 23-54). Cambridge: Massachusetts Institute of Technology.

Gaukroger, S. (1998/2004). Introduction. En S. Gaukroger (Ed.). *Descartes: The world and other writings* (pp. 248). Nueva York: Cambridge University Press.

Griffiths, P. (1997). *What emotions really are*. Chicago: The University of Chicago Press.

Hacker, P. M. S. (1972/1986). Wittgenstein's later conception of philosophy. *Insight and illusion: themes in the philosophy of Wittgenstein* (pp. 146-178). Nueva York: Oxford University Press.

Hacker, P. M. S. (2004/2009). The conceptual framework for the investigation of the emotions. En Y. Gustafsson, C. Kronqvist, & M. McEachrane (Eds.), *Emotions and understanding: Wittgensteinian perspectives* (pp. 43-59). Nueva York: Palgrave Macmillan.

Hatfield, G. (2007a). Did Descartes have a Jamesian theory of the emotions? *Philosophical Psychology, 20*(4), 413-440.

Hatfield, G. (2007b). The passions of the soul and Descartes's machine psychology. *Studies in History and Philosophy of Science*, Part A, *38*(1), 1-35.

Helm, B. W. (2009). Emotions as evaluative feelings. *Emotion Review, 1*(3), 248-255.

Horst, S. (2007). *Beyond reduction: philosophy of mind and post-reductionist philosophy of science*. Nueva York: Oxford University Press.

James, W. (1884). What is an emotion? *Mind,* (34), 188-205.

James, W. (1890/1950). *The principles of psychology*. Nueva York: Dover Publications.

James, W. (1892). A plea for psychology as a 'natural science'. *The Philosophical Review, 1*(2), 146-153.

James, W. (1892/2001). *Psychology: the briefer course*. Mineola: Dover Publications.

James, W. (1894/1994). The physical basis of emotion. *Psychological Review, 101*(2), 205-210.

Kenny, A. (1963/2003). *Action, emotion and will*. Nueva York: Routledge.

Lakatos, I. (1978/2010a). La falsación y la metodología de los programas de investigación científica (J. C. Zapatero, Trad.). *La metodología de los programas de investigación científica*. Madrid: Alianza.

Lakatos, I. (1978/2010b). La historia de la ciencia y sus reconstrucciones racionales (J. C. Zapatero, Trad.). *La metodología de los programas de investigación científica*. Madrid: Alianza.

Lazerowitz, M. (1968). Wittgenstein: the nature of philosophy. *Philosophy and Illusion* (pp. 53-70). Nueva York: Humanities Press.

Moya, C. J. (1994). Las emociones y la naturalización de la intencionalidad. *Logos. Anales del Seminario de Metafísica*, (28), 227-255.

Nussbaum, M. C. (2001). Emotions as judgements of value *Upheavals of thought: the intelligence of emotions* (pp. xiii, 751 p.). Nueva York: Cambridge University Press.

Palencik, J. T. (2007). William James and the psychology of emotion: From 1884 to the present. *Transactions of the Charles S Peirce Society, 43*(4), 769-786.

Prinz, J. J. (2004a). Embodied emotions. En R. C. Solomon (Ed.). *Thinking about feeling: Contemporary philosophers on emotions* (pp. 44-60). Nueva York: Oxford University Press.

Prinz, J. J. (2004b). *Gut reactions: a perceptual theory of emotion*. Nueva York: Oxford University Press.

Putnam, H. (1981/2006). *Razón, verdad e historia* (J. M. Esteban Cloquell, Trad.). Madrid: Tecnos.

Quine, W. V. O. (1960). *Word and object*. Cambridge: MIT Press.

Ratcliffe, Matthew (2005). William James on emotion and intentionality. *International Journal of Philosophical Studies* 13(2), 179–202.

Shapiro, L. (2003). Descartes passions of the soul and the union of mind and body. *Archiv für Geschichte der Philosophie, 85*(3), 211-248.

Solomon, R. C. (1973/1980). Emotions and choice. En A. Oksemberg Rorty (Ed.), *Explaining emotions* (pp. 251-282). Los Ángeles: University of California Press.

Wilshire, B. (1969). Protophenomenology in the psychology of William James. *Transactions of the Charles S. Peirce Society, 5*(1), 25.

Wittgenstein, L. (1921/1999). *Tractatus logico-philosophicus* (Reguera Pérez, I. & Muñoz Veiga, J., Trads.). Madrid: Alianza.

Wittgenstein, L. (1953/2003). *Investigaciones filosóficas* (A. G. Suárez & U. Moulines, Trads.). México: Universidad Nacional Autónoma de México.

Wittgenstein, L. (1958/2009). *Los cuadernos azul y marrón* (F. G. Guillén, Trad.). Madrid: Tecnos.

Wittgenstein, L. (1966/1992). *Lecciones y conversaciones sobre estética, psicología y creencia religiosa* (I. Reguera, Trad.). Barcelona: Paidós Ibérica.

Wittgenstein, L. (1969/2006). *Sobre la certeza* (J. L. Prades & V. Raga, Trads.). Barcelona: Gedisa.

Wittgenstein, L. (1980/2007). *Observaciones sobre la filosofia de la psicologia* (L. F. Segura, Trad., G. E. M. Anscombe & G. H. von Wright Eds. 2ª ed.). México: Universidad Nacional Autónoma de México.

Woodruff Smith, D. (1999/2004). Intentionality naturalized? *Mind World. Essays in Phenomenology and Ontology* (p. 176-210). Nueva York: Cambridge University Press.

Este libro fue compuesto en caracteres
Stempel Garamond 12 puntos, sobre
papel propal de 70 gramos y encuadernado
con método *Hot Melt*, en mayo de 2016,
en Bogotá, D. C., Colombia
Digiprint